In *The evolution of intelligence*
R. J. Sternberg & J. C. Kaufman (Eds.),
Hillsdale, NJ: Erlbaum.

8

Unraveling the Enigma of Human Intelligence: Evolutionary Psychology and the Multimodular Mind

Leda Cosmides
John Tooby
University of California

EVOLUTIONARY PSYCHOLOGY AND THE ENIGMA OF INTELLIGENCE

Evolution brought brains and minds into a world initially devoid of intelligent life. The evolutionary process designed the neural machinery that generates intelligent behavior, and important insights into how this machinery works can be gained by understanding how evolution constructs organisms. This is the rationale that underlies research in evolutionary psychology.

Evolutionary psychology was founded on interlocking contributions from evolutionary biology, cognitive science, psychology, anthropology, and neuroscience. It reflects an attempt to think through, from first principles, how current knowledge from these various fields can be integrated into a single, consistent, scientific framework for the study of the mind and brain (Cosmides & Tooby, 1987; Pinker, 1997; Tooby & Cosmides, 1992b).

Perhaps more than any other issue, questions about the nature and evolution of human intelligence and rationality have played a central organizing role in the development of evolutionary psychology. Indeed, how evolutionary psychologists answer questions about the evolutionary basis of intelligence demarcates it from more traditional behavioral science approaches. As a starting point, evolutionary psychologists share with other cognitive scientists a commitment to discovering exactly how mental operations are realized computationally and physically in the mind and brain. To this, they add a perspective that attempts to incorporate knowledge about the brains and natural behavior of each species that has been studied, and a recognition that the evolutionary process constructed the computational systems present in the minds of organisms primarily through natural selection (Cosmides & Tooby, 1987; Pinker, 1997; Tooby & Cosmides, 1992b).

To make progress in understanding the phenomenon of evolved intelligence, we have been led to distinguish two related meanings of intelligence. We call these *dedicated intelligence* and *improvisational intelligence*. Dedicated intelligence refers to the ability of a computational system to solve a predefined, target set of problems. Improvisational intelligence refers to the ability of a computational system to improvise solutions to novel problems. Ordinary use of the term *intelligence* is inconsistent: People sometimes use it to mean something similar to improvisational intelligence. But the term is also often applied to systems that are highly successful at solving their respective problems, regardless of whether the problem is novel or the solution improvised. People remark on the intelligence of such things as the bat's sonar navigation system, more accurate bombs, the rice cooker with sensors and fuzzy logic circuits that decide when the rice is done, and Sojourner, the semiautonomous rover that explored the surface of Mars. Distinguishing between these two types of intelligence is indispensable for understanding how evolution constructed intelligent circuitry in organisms.

Traditionally, many behavioral and social scientists have, implicitly or explicitly, believed the following:

1. Humans are endowed with improvisional intelligence.
2. Most human behavior is explained by the operation of improvisional intelligence.
3. Most of our interesting and important psychological operations are the output of a system for improvisional intelligence.
4. Improvisional intelligence is achieved by an architecture that is essentially a blank slate connected to general-purpose (content-independent, domain-general) reasoning and learning circuits.
5. Improvisional intelligence is easy, at least in concept, to understand and to design, and might soon be built into artificial systems.

6. Specialized programs, because they are inflexible, would hamper or reduce the intelligence of a system.
7. Therefore, humans evolved intelligence by giving up instincts and innate structure and substituting general-purpose learning, reasoning, and intelligence instead.

In contrast, we have come to the following conclusions:

1. Humans are endowed with improvisational intelligence, but
2. Humans are also endowed with a large and heterogeneous set of evolved, reliably developing, dedicated problem-solving programs, each of which is specialized to solve a particular domain or class of adaptive problems (e.g., grammar acquisition, mate acquisition, food aversion, way-finding).
3. Each such neural program exhibits a well-engineered, problem-solving intelligence when applied to the targeted set of problems it evolved to solve. However, these adaptive specializations cannot, by their nature, display improvisional intelligence, at least individually.
4. A very large proportion of human thought and action owes its intelligent patterning to dedicated problem-solving intelligence rather than improvisional intelligence.
5. The larger the number of dedicated intelligences a system has, the broader the range of problems it could solve.
6. For reasons rooted in the nature of computation and in the way natural selection works, improvisional intelligence is difficult to implement and to evolve, and presents deep theoretical challenges. In short, the puzzle of how improvisional intelligence is computationally and evolutionary possible is a profound one.
7. Nevertheless, improvisional intelligence might have been achieved through (a) bundling an increasing number of specialized intelligences together and (b) embedding them in an encompassing architecture that has a *scope syntax*: an elaborate set of computational adaptations for regulating the interaction of transient and contingent information sets within a multimodular mind.

In short, evolutionary psychologists have arrived at a series of sometimes heterodox conclusions about what intelligence means, how it is constructed, and what role intelligence plays in the human psychological architecture. The remainder of the chapter sketches out the logic that has led to these conclusions (see also Cosmides & Tooby, 1987, in press; Tooby & Cosmides, 1992a, 1992b; Tooby & DeVore, 1987). In order to retrace these steps, we will need to address a series of fundamental questions, such as What is intelligence? What is computation? and What is an adaptive problem?

The Robot Challenge

The fields of cognitive psychology and artificial intelligence grew up together, and their animating questions became deeply intertwined. The pioneering work of mathematicians and early computer scientists, such as Alan Turing, John Von Neuman, Alan Newell, and Herbert Simon, set off a race to create intelligent machines, where intelligence was defined with respect to a cultural standard of general problem solving. The goal of developing a causal account of how thought can be produced by a mechanical system was shared by both cognitive psychologists and computer scientists. As many thought of it, the primary difference between the two fields was whether the mechanical system in question was a carbon-based brain or a silicon-based computer, and researchers debated whether this difference in physical substrate was trivial or would constrain, in important ways, the kinds of computations that each system could perform. In this atmosphere, many discussions of intelligence were framed by what one can think of as the robot challenge: What criteria would a robot have to meet before it was said to exhibit humanlike intelligence? What programs would the robot need in order to achieve these criteria?

Steven Pinker (1997) formulated one of the clearest analyses of the robot challenge. In Pinker's view, intelligence is "the ability to attain goals in the face of obstacles by means of decisions based on rational (truth-obeying) rules" (p. 62), where *rational* and *truth-obeying* are understood not in the narrow logician's sense but in the broader sense of rules that correspond to reality, at least in the statistical sense. In arguing for this definition, he points out that (i) without a specification of a creature's goals, the concept of intelligence is meaningless (is a toadstool brilliant because it is good at remaining exactly where it is?); (ii) we would be hard pressed to credit an organism with much intelligence if, in attempting to overcome obstacles to achieve goals, its actions were unconnected to reality (e.g., wanting to split a log, it hacks at empty space); and (iii) overcoming obstacles implies the ability to shift to different plans of action, depending on the nature of the obstacle. Different means are chosen to achieve the same end, depending on the particulars of the situation one is facing. According to Pinker, any system exhibiting this property—robot, space alien, or earth species—would count as having "rational, humanlike thought." Pinker's definition elegantly captures many intuitions that people have about intelligence—at least of the human variety—and provides a clear foundation for thinking about the question. It also encapsulates much of what we mean when we speak of improvisional intelligence.

Indeed, views such as this have organized the thinking of philosophers and scientists for many centuries. What kind of mental machinery does an organism need to manifest this form of intelligence? Evolutionary psychologists argue

that there are actually many different possible answers to this question (Cosmides & Tooby, in press; Pinker, 1994, 1997; Tooby & Cosmides, 1992b). However, this is not the traditional view. Noting that humans—unlike many other animals—are able to pursue so many different goals, overcoming so many different obstacles using so many different means, many thinkers have assumed that the nature of the mental machinery that creates intelligence in humans must be free of anything that might constrain it; that is, it must be a blank slate. The flexibility of human intelligence—that is, our ability to solve many different kinds of problems—was thought to be conclusive evidence that the circuits that generate it are general purpose and content free. *Homo sapiens* was thought of as the one animal endowed with reason, a species whose instincts were erased by evolution because they were rendered unnecessary by (or were incompatible with) culture, the ability to learn, and intelligence. This conception of the nature of human intelligence has been a central pillar of what we have called the standard social science model (SSSM), that is, the worldview that has dominated the social and behavioral sciences for the past century (for an extended dissection of this paradigm, see Tooby & Cosmides, 1992b).

The Standard Social Science Model

The SSSM maintains that the human mind is a blank slate, virtually free of content until written on by the hand of experience. According to the 13th-century philosopher Aquinas, there is "nothing in the intellect that was not previously in the senses." Working within this framework, the 17th- and 18th-century British Empiricists and their successors produced elaborate theories about how experience, refracted through a small handful of innate mental procedures, inscribed content onto the mental slate.

Over the years, the technological metaphor used to describe the structure of the human mind has been consistently updated, from blank slate to switchboard to general purpose computer. But the central tenet of these Empiricist views has remained the same: All of the specific content of the human mind originally derives from the "outside"—from the environment and the social world—and the evolved architecture of the mind consists solely or predominantly of a small number of general purpose mechanisms that are content-independent and that are referred to using terms such as *intelligence*, *learning*, *induction*, *imitation*, *rationality*, and *the capacity for culture*.

So according to this view, the same mechanisms are thought to govern how one acquires a language, learns to recognize emotional expressions, responds to the possibility of incest, responds to an attack or flattery, or adopts ideas about friendship and reciprocity (indeed everything but perception, which is often accepted as being specialized and at least partly innately structured). The mechanisms that govern reasoning, learning, and memory are hypothesized to operate *uniformly, according to unchanging principles,* regardless of the content they are operating on or the larger category or domain involved. (For this reason, we call such hypothesized mechanisms *content-independent* or *domain-general.*) Such mechanisms, by definition, have no preexisting content built in to their procedures; they are not designed to construct certain contents more readily than others; and they have no features specialized for processing particular kinds of content more than others. Because these hypothetical mental mechanisms have no content of their own to impart, it logically follows that all the particulars of what we think and feel are derived externally, from the physical and social world. In this view, the evolutionary process explains the evolution of the human body, human intelligence, and the capacity for learning culture, but the blank slate nature of the human mind interposes a barrier between biology and human mental content that renders evolution essentially irrelevant to human affairs. Unlike other animals, our evolution washed us clean of instincts and innate mental organization. So, the issue of the nature of human intelligence, and the role that it plays in the operation of the human mind, is not a minor one. Beliefs about intelligence ramify far beyond psychology, into every aspect of the behavioral and social sciences. Although there have been intense controversies about the significance of individual differences in intelligence and its measurement, its larger theoretical role as the central concept explaining how humans differ from other species, acquire culture, and generate the majority of their behavior has seemed almost self-evident to scholars.

Nevertheless, we think that three decades of converging research in cognitive psychology, evolutionary biology, anthropology, and neuroscience have shown that this plausible and persuasive view of the human mind is incorrect. Evolutionary psychology represents an alternative proposal about how to organize our understanding of the human mind and the nature of human intelligence. According to this alternative perspective, all normal human minds reliably develop a standard collection of reasoning, emotional, and motivational circuits or programs. These programs were functionally designed over evolutionary time by natural selection acting on our hunter–gatherer (and more distant) ancestors. They are individually tailored to the demands of particular evolutionary functions and often come equipped with what philosophers would once have called "innate ideas." There are far more of them than anyone had suspected, and they respond far more sensitively to the particulars of human life than anyone had imagined. Humans appear to have evolved circuits specialized for the domains of friendship, incest avoidance, coalitions, landscape preference, status, number, aggression, mating, language, intuiting what others are thinking, judging personality, and hundreds of other functions. These circuits organize the way we interpret our experiences, inject certain recurrent concepts and motivations into our mental life, give us our passions, and provide cross-culturally universal frames of meaning that allow us to understand the actions and intentions of others and to acquire the locally variable components of culture (for relevant reviews, see, e.g., Barkow, Cosmides, & Tooby, 1992; Gallistel, 1990; Hirchfeld & Gelman, 1994; Pinker, 1994, 1997).

The Organismic Challenge

The robot challenge grew out of the concerns of cognitive scientists interested in machine intelligence. But we would like to propose two definitions of intelligence that grow out of the concerns of evolutionary biologists, behavioral ecologists, and others who study animal behavior. The world is full of millions of species, all of whom have succeeded in surviving and reproducing in a world of fierce antagonists, entropy, and harsh environmental reverses. No existing robot or computer comes close to solving the stringent problems routinely faced by members of the living world. Facts about the living world constitute the organismic challenge: What criteria would an organism have to meet before it was said to exhibit some form of intelligence? What kind of programs would the organism need in order to achieve these criteria?

As special as human intelligence may be—and we do believe that it is zoologically unprecedented—one does see other animals overcome obstacles to attain goals, and their decisions take into account real facts about the world. The goals pursued may be different from ours; the range of possible goals pursued by members of any one species may be more limited, and the variety of means any one species employs in attaining them may be more limited as well. Nevertheless, everyone recognizes that the animals that surround us routinely overcome obstacles to attain goals, even if (to nonbiologists) the status of other organisms, such as plants, fungi, protists, and prokaryotes, is less clear.

Although nonbiologists are frequently unaware of the subtlety, intricacy, elegance, and sophistication expressed in the behavior of nonhumans, there is now a wealth of data available that needs to be assimilated into a general consideration of natural intelligence. Over the last 30 years, there has been an explosion of research in field biology, the rapid development of new experimental methods, and dramatic advances in adaptationist evolutionary biology that together provide a panorama of superb computational problem solving applied to a immense array of adaptive problems by a multiciplicity of species. For example, having wandered far in search of food in terrain that is often devoid of landmarks, desert ants return home, directly, by a straight line route, a feat they accomplish through vector integration (Gallistel, 1990; Wehner & Srinivasan, 1981). During classical conditioning, pigeons, rats, and other animals perform computations that are equivalent to a nonstationary multivariate time series analysis: From noisy, changing data, they figure out the contingencies between events in the world (Gallistel, 1990). Migratory birds extract configural relationships from the constellations and use them to navigate across thousands of miles. Rats, which evolved to be opportunistic omnivores, have such sophisticated strategies for testing novel foods that they routinely outwit exterminators attempting to poison them (Kalat & Rozin, 1973; Rozin & Kalat, 1971). Zebras continue to feed if they detect that the nearby lion has fed recently, and bother to interrupt their feeding only if they have insufficient evidence that this is the case. Male mice often kill unrelated baby mice, an act that causes the dead infants' mothers to reach estrus far earlier than they would if they had continued to nurse the unrelated male's offspring. They do not, however, kill their own pups: A male's first intravaginal ejaculation starts a neural timer that counts off the days of a typical pregnancy, and he stops committing infanticide several days before the birth of babies that could, in principle, be his own (Perrigo, Bryant, & vom Saal, 1990). A male dunnock will feed the chicks of the female he has been mating with in proportion to the probability that her babies are his as opposed to the coresident male's (Burke, Davies, Bruford, & Hatchwell, 1989; Davies, 1989). A stickleback fish will risk his life in defense of his nestful of eggs in proportion to the number of eggs in it (Pressley, 1981). Desert rodents manage their larder of seeds, taking into account the age of the seeds, their stage of germination, their nutritional value, the humidity in each area of the cache, the cost of acquisition, and many other variables (Gendron & Reichman, 1995; Post, McDonald, & Reichman, 1998). Chimpanzees engage in Machiavellian political intrigues, involving shifting coalitions and alliances (de Waal, 1982). In all these cases, the animals are using information about changes in the state of the world or the value of a resource to adjust their behavior in ways that achieve adaptive outcomes.

WHAT IS INTELLIGENCE?

To analyze these forms of intelligence, which are so abundantly manifest in the animal world, and to explore how they might relate to the emergence of human intelligence, it is necessary to introduce two definitions that distinguish two meanings of intelligence that apply to organisms. Because these forms of intelligence arose naturally, through the process of evolution, we think a number of insights might come from grounding the analysis of intelligence within the causal framework of evolutionary biology. For one thing, there are constraints on what kinds of machinery natural selection can design, and this will affect the form that intelligence takes. In particular, developing a conception of intelligence that can be applied widely to organisms allows us to zero in on those aspects of human intelligence that may be zoologically unique. Therefore, we would like to define two forms of intelligence as follows:

Intelligence$_1$. A computational system or program is intelligent$_1$ when it is well designed for solving a target set of adaptive computational problems. We will call this *dedicated intelligence*.

Intelligence$_2$. A computational system is intelligent$_2$ to the extent that it is well designed for solving adaptive computational problems, and has components designed to exploit transient or novel local conditions to achieve adaptive outcomes. We will call this *improvisational intelligence*.

To understand what these definitions mean, we need to say more precisely what we mean by *computational*, *designed*, *adaptive problem*, *adaptive outcome*, *transient*, *novel*, and *local*. These terms are defined with respect to one another, within a causal framework provided by Darwin's theory of evolution by natural selection.

What Is a Computational System?

Organisms are composed of many parts. Some of these parts are computational. By computational, we mean that they are designed to (i) monitor the environment for specific changes and (ii) regulate the operation of other parts of the system functionally on the basis of the changes detected. For example, the diaphragm muscle, which causes the lungs to contract and expand, is not computational. But the system that measures carbon dioxide in the blood and regulates the contraction and extension of the diaphragm muscle is. The plastic cover on a thermostat is not computational, nor are the parts of a furnace that generate heat. But the thermocouple that responds to ambient temperature by toggling the switch on the furnace, and the connections between them, form a computational system. Muscles are not computational, but the visual system that detects the presence of a hungry-looking lion, the inference mechanisms that judge whether that lion has seen you or not, and the circuits that cause your muscles to either run to a nearby tree (if the lion has seen you) or freeze (if it hasn't seen you) do compose a computational system. The language of information processing can be used to express the same distinction: One can identify the computational components of a system by isolating those aspects that were designed to regulate the operation of other parts of the system on the basis of *information* from the internal and external environment.

By "monitoring the environment for specific changes," we mean the system is designed to detect a change in the world. That change can be internal to the organism (such as fluctuations in carbon dioxide levels in the blood or the activation of a memory trace) or external to the organism (such as the onset of a rainstorm or the arrival of a potential mate). Changes in the world become *information* when (i) they interact with a physical device that is designed to change its state in response to variations in the world (i.e., a transducer), and

(ii) the changes that are registered then participate in a causal chain that was designed to regulate the operation of other parts of the system. A photon, for example, does not become information until it causes a chemical reaction in a retinal cell, which was designed for this purpose and is part of a causal system that was itself designed to regulate an organism's behavior on the basis of inferences about what objects exist in the world and where they are.

A set of features is not computational unless they were *designed* to exhibit these properties. For example, the outer cells of a dead tree stump expand in the rain, and as this happens, the inner portions of the stump might become compressed. But these dead cells were not designed for detecting changes in weather. More important, although their swelling does cause a change in the inner part of the stump, it is not *regulating* the operation of the stump. Regulation means more than merely influencing or changing something. It means systematically modifying the operation of a system so that a *functional* outcome is achieved. In the case of a thermostat, that function was determined by the intentions of the engineer who designed it. In the case of an organism, that function was determined by natural selection, which acted to organize the properties of the organism.

A causal process does not need the human properties of foresight and intention to be capable of designing something. The selection of parts on the basis of their functional consequences is the crux of the concept of design (e.g., we say a thermocouple has been designed because the two different metals, each with different heat-conducting properties, did not come together by chance; they were selected for the thermocouple *because* this has functional consequences if one's purpose is to regulate something's temperature). From this perspective, it does not matter whether the causal system that does the selection is a volitional agent or a feedback process. A system can be said to be designed whenever the cause of its having the parts and properties that it has—rather than others—is that they have functional consequences, i.e., that they solve a problem of some kind (see, e.g., Nozick, 1993, p. 118). By this criterion, natural selection designs organisms. Chance events, such as mutations, cause alternative parts (design features) to be introduced into a population of organisms, but natural selection is not a chance process. Natural selection is a systematic feedback process that retains or discards parts because of their consequences on the functional performance of the system.

How Natural Selection Designs Organisms

The heart of Darwin's insight is the recognition that organisms are self-reproducing machines (Dawkins, 1976, 1986; Williams, 1966). From a Darwinian perspective, the defining property of life is *reproduction*, or more fully, the presence in a system of devices (organized components) that cause the system to construct new and similarly reproducing systems. These organized components can be thought of as design features: They are present because they participate in the causal process whereby the organism produces new organisms with a similar structure, (i.e., with a similar design). One can consider design features at many scales from, for example, the visual system, the eye, and the retina, down to the retinal cells, their organelles, and the photoreactive pigments that trigger the firing of the cell.

Individuals die, but their design features live on in their descendants—if they have any. When an organism reproduces, replicas of its design features are introduced into its offspring. But the replication of the design of the parental machine is not always error free. As a result, randomly modified designs (i.e., mutants) are introduced into populations of reproducers. Because living machines are already exactingly organized so that they cause the otherwise improbable outcome of constructing offspring machines, random modifications will usually introduce disruptions into the complex sequence of actions necessary for self-reproduction. Consequently, most newly modified but now defective designs will remove themselves from the population—a case of negative feedback.

However, a small residual subset of design modifications will, by chance, happen to constitute improvements in the system's machinery for causing its own reproduction. Such improved designs (by definition) cause their own increasing frequency in the population—a case of positive feedback. This increase continues until (usually) such modified designs outreproduce and thereby replace all alternative designs in the population, leading to a new species-standard design. After such an event, the population of reproducing machines is different from the ancestral population: The population- or species-standard design has taken a step uphill toward a greater degree of functional organization for reproduction than it had previously. Over the long run, down chains of descent, this feedback cycle pushes designs through state-space toward increasingly well-engineered—and otherwise improbable—functional arrangements. These arrangements are functional in a specific sense: The elements are well organized to cause their own reproduction in the environment in which the species evolved.

For example, if a more sensitive retina, which appeared in one or a few individuals by chance mutation, causes predators to be detected more quickly, individuals who have the more sensitive retina will produce offspring at a higher rate than those who lack it. Those of their offspring that inherit that more sensitive retina will also evade predators better and therefore produce offspring at a higher rate, and so on down the generations. By promoting the reproduction of its bearers, the more sensitive retina thereby promotes its own spread over the generations, until it eventually replaces the earlier model retina and becomes a universal feature of that species' design. This spontaneous feedback process—natural selection—causes functional organization to emerge naturally and inevitably, without the intervention of an intelligent designer or supernatural forces. Genes are simply the means by which design features replicate themselves from parent to offspring. They can be thought of as particles of design: elements that can be transmitted from parent to offspring and that, together with an environment, cause the organism to develop some design features and not others. Because design features are embodied in individual organisms, there are usually only two ways they can propagate themselves: by solving problems that increase the probability that offspring will be produced by either the organism they are situated in, or by that organism's kin. An individual's relatives, by virtue of having received some of the same genes from a recent common ancestor, have an increased likelihood of having the same design feature as compared to other conspecifics. This means that a design feature in an individual that causes an increase in the reproductive rate of that individual's kin will, by so doing, tend to increase its own frequency in the population. A computational element that causes an individual to be motivated to feed her sisters and brothers, if they are starving, is an example of a design circuit that increases kin reproduction. When the individual's siblings reproduce, they might pass on this same circuit to their children. Hence, design features that promote both direct reproduction and kin reproduction, and that make efficient trade-offs between the two, will replace those that do not. How well a design feature systematically promotes direct and kin reproduction is the bizarre but real engineering criterion determining whether a specific design feature will be added to or discarded from a species' design. Therefore, we can potentially understand why our brains are constructed in the way they are, rather than in other perfectly possible ways, when we see how its circuits were designed to cause behavior that, in the world of our ancestors, led to direct reproduction or kin reproduction.

Computational and Noncomputational Adaptive Problems

We can now define the concept of adaptive behavior with precision. Adaptive behavior, in the evolutionary sense, is behavior that tends to promote the reproduction of the design feature into the next generation (which usually means increasing the net lifetime reproduction of an individual bearing the design feature or that individual's genetic relatives). By promoting the replication of the genes that built them, circuits that—systematically and over many generations—cause adaptive behavior become incorporated into a species' neural design. In contrast, behavior that undermines the reproduction of the individual or his or her blood relatives removes the circuits causing those behaviors from the species, by removing the genes that built those circuits. Such behavior is *maladaptive*, in the evolutionary sense.

So, evolutionists continually analyze how design features are organized to contribute to lifetime reproduction, not because of an unseemly preoccupation with sex, but because reproduction was the final causal pathway through which a functionally improved design feature caused itself to become more numerous with each passing generation, until it became standard equipment in all ordinary members of the species.

Enduring conditions in the world, such as the presence of predators, the need to share food to buffer against bad luck in food acquisition, or the vulnerability of infants, constitute *adaptive problems*. Adaptive problems have two defining characteristics. First, they are conditions or cause-and-effect relationships that many or most individual ancestors encountered, reappearing again and again during the evolutionary history of the species. Second, they are problems whose solution increased the reproduction of individual organisms or their relatives—however indirect the causal chain, and even if the effect on the organism's own offspring or the offspring of kin was relatively small. Most adaptive problems have to do with relatively mundane aspects of how an organism lives from day to day: what it eats, what eats it, who it mates with, who it socializes with, how it communicates, and so on.

A subset of adaptive problems are computational. Adaptive computational problems are those problems that can be solved by design features that monitor some aspect of the environment (either internal or external) and use the information detected to regulate the operation of other parts of the organism. Those parts of an organism that were designed to regulate its behavior on the basis of information are computational. To say these parts were designed for this purpose means that their contribution to this regulatory process was one of the functional consequences that caused them to be incorporated into the species's architecture by natural selection. There can, of course, be computational systems that regulate the operation of subsystems that are not behavioral, at least in the colloquial sense (e.g., the system in a mother that detects how much an infant is sucking at the breast and adjusts milk production on the basis of this information would be a computational system).

What Does *Well-Designed* Mean?

An enduring adaptive problem constantly selects for design features that promote the solution to that problem. Over evolutionary time, more and more design features accumulate that fit together to form an integrated structure or device that is well engineered to solve its particular adaptive problem. Such a structure or device is called an *adaptation*. Indeed, an organism can be thought of as largely a collection of adaptations, such as the functional subcomponents of the eye, liver, hand, uterus, or circulatory system. Each of these adaptations exists in a species' design now because it contributed ancestrally to the process of self and kin reproduction.

So natural selection builds adaptations—that is, problem-solving machinery—to solve evolutionarily long-standing adaptive problems, and some of these problems are computational in nature. One can identify an aspect of an organism's physical or psychological structure—its phenotype—as an adaptation by showing that (i) it has many design features that are improbably well suited to solving an ancestral adaptive problem, (ii) these phenotypic properties are unlikely to have arisen by chance alone, and (iii) they are not better explained as the by-product of mechanisms designed to solve some alternative adaptive problem or some more inclusive class of adaptive problem. Finding that an architectural element solves an adaptive problem with reliability, precision, efficiency, and economy is prima facie evidence that one has located an adaptation (Williams, 1966). Ultimately, the objective measure of engineering quality is how much better than random a system is at meeting its functional goals. Intuitively, however, we can appreciate the quality of evolved systems by comparing them, where feasible, to human efforts.

Using this standard, evolved systems are not optimal or perfect (whatever that may mean), but they are very good by human engineering standards. We can say this with confidence because human engineers—even when they have enormous research budgets and can devote decades to a single project—have not been able to match the quality of what evolution produces. Skeptics of the power of natural selection have based their skepticism on verbal assertion rather than any comparison of the performance of human-engineered and evolutionarily engineered systems (e.g., Gould & Lewontin, 1979). Natural selection has produced exquisitely engineered biological machines—grammar acquisition, object recognition, word-meaning induction, the regulation of walking, tactile perception, olfaction, color constancy systems, solar energy capture—whose performance is unrivaled by any machine yet designed by humans.

WHAT IS *DEDICATED INTELLIGENCE*?

It should now be clear what we mean by our proposed definition of targeted intelligence, as applied to organisms: A neural program manifests *dedicated intelligence* when it is well designed for solving a targeted set of adaptive computational problems. (This is similar to the concept of ecological rationality developed in Tooby & Cosmides, 1992b).

Researchers know about thousands of systems of dedicated intelligence in humans and other species, designed for the regulation of food choice, mate choice, alliance maintenance, predator-escape, contagion avoidance, thermoregulation, fluid intake, social status, sex changes, aphid farming, landmark recognition, grammar acquisition, child survival, deception detection, aggression, patch selection in foraging, incest avoidance, dead-reckoning, coalition formation, offspring recognition, birth regulation, sex ratio manipulation, fungus-growing, web-building, blood pressure management, celestial navigation, competitive infanticide, snake avoidance, toxin assessment, and everything else necessary to maintain the innumerable alternative ways of life exhibited by earth's species.

By intention, this definition of dedicated intelligence is agnostic on several issues. For example:

- *It does not rest on any specific conception of the nature of the computational machinery that produces solutions to adaptive problems.* Natural selection has come up with an immense diversity of solutions to various adaptive problems, and there are no grounds for prejudging the methods by which adaptive computational problems might be solved (Tooby & Cosmides, in press). This contrasts with some approaches to assessing problem-solving performance (e.g., Kahneman, Slovic, & Tversky, 1982). In the judgment and decision-making community, for example, researchers often define a subject as rational only if he or she adheres to the experimenter's preferred procedure for decision making, where the procedure is usually some formalism derived from mathematics, probability theory, or logic, such as Bayes' Rule, or *modus tollens*. This is like grading sharpshooters on the basis of their form in holding the rifle, instead of on how often they hit the target.
- *It does not depend on the presence of a brain.* By this definition, there can be intelligent systems distributed throughout an organism's body, which is fortunate, because all bodies contain highly sophisticated computational regulatory processes. They need not all be localized with one another in a central computational organ. Thus, this definition includes organisms equipped with distributed cognition (e.g., decentralized systems composed of sensors and springs in the limbs of an organism that adjust their motion sensitively to details of the terrain, Clark, 1997). Indeed, phylogenetically, distributed intelligence undoubtedly appeared before it became concentrated into brains. It would be arbitrary to tie the definition of intelligence to the distribution of its physical basis rather than to its regulatory performance.
- *It does not depend on the existence of a mentally represented goal.* Not all behavior that looks goal directed involves representations of goals. For example, ticks have a circuit directly linking chemoreceptors to motor neurons, so that the smell of butyric acid causes the tick to drop from a tree (Uexkull, 1905/1957). Because butyric acid is emitted only by mam-

mals, this circuit usually results in the tick landing on a mammalian host, whose blood it then drinks. The design of this circuit makes the tick's behavior appear goal directed. Yet it involves no explicit representation of a goal state. Nevertheless, this computational system clearly exhibits dedicated intelligence. The simplicity of this system is not the issue. Computation without explicit goals can involve any level of complexity. For example, it seems unlikely that either vision or spontaneous early grammar acquisition involve explicitly represented goals, but both involve very intricate computational processes. So, a system can exhibit targeted intelligence whether or not it explicitly represents goal states, and humans appear to have intelligent programs of both kinds. Of course, explicitly represented goal states are a necessary feature of improvisional intelligence, as we discuss.

- *The requirement that intelligent machinery be adaptively well designed introduces criteria such as economy, efficiency, precision, and reliability into the analysis of intelligence.* Not only is there a biological justification for this, but this matches our intuitions as well. Consider two desert ants, equipped with two different navigational designs, facing the problem of returning to the nest. One travels the shortest distance, thereby saving energy and reducing the amount of time she spends above ground, where she is at risk of being predated upon. The other meanders across the landscape without doing anything functional on this longer path, although she also eventually reaches home. Biologically, one design is better than the other (because it has solved the problem more efficiently), which parallels our intuition that the first ant has behaved more intelligently than the second. The cost of running the computational system is also part of the analysis, and so the ultimate currency for comparing alternative designs for fulfilling the same function is the net fitness produced over the set of conditions being considered.
- *The definition refers to adaptations, programs, or systems, not to entire organisms.* It provides criteria for judging whether any particular subsystem exhibits dedicated intelligence but cannot be used to assess the intelligence of the organism as a whole using a single-dimensional variable. For example, a bee has foraging algorithms that are very well engineered for foraging among flowers (Heinrich, 1981), but it lacks the ability to navigate by the stars or (we suspect) to create or track false beliefs in social competitors. Similarly, strokes can knock out a person's ability to speak grammatically, yet leave intact their ability to think spatially. Both species and individual organisms will embody distinct complexes of specific abilities. Therefore, this definition is incompatible with a framework that necessarily views intelligence as a unitary phenomenon and attempts to array species along a continuum of

more or less intelligent. Also, by applying the criterion of how well designed a computational system is at solving a particular class of adaptive problem, this definition does not prejudge whether an organism's improvisional intelligence is achieved via a bundle of dedicated computational modules or by a single, general purpose system.
- *This definition distinguishes between the design of a system and the outcome achieved by a particular organism in a particular instance.* One cannot judge intelligence or how well engineered a dedicated computational system is by its performance on a single occasion for the same reason that the value of a betting system cannot be evaluated by what happens on a single bet. The quality of the dedicated intelligence in a computational system is a function of the performance of the system summed across the range of environments considered relevant to the evaluation. For natural selection, the range of relevant environments is the distribution of conditions faced by the ancestors of the species during their evolution. For example, savannah predators often ambushed their prey from trees. A well-designed computational system that evolved to function in that environment might routinely cause prey to spend a few extra calories walking around a tree that is too dense to see through, even though in 999 out of 1,000 cases the tree is predator free. Descendants of such prey, such as humans, might still find visually impenetrable, overhead foliage mildly disquieting, even in a postindustrial world where there are no longer leopards or saber-toothed tigers. Nevertheless, that computational system is still manifesting intelligence$_1$.
- *The degree of dedicated intelligence displayed by a neural program is relative to the ecological structure of the world the organism inhabits and to the problem-solving goals posed by its associated adaptive problem.* Once a target set of outcomes is specified (what behaviors solve the adaptive problem), any number of alternative computational designs can be compared by examining how well each performs in reaching the goal. The better a design is at reaching the goal, the more dedicated intelligence it shows. On this view, the intelligence of a program design would consist of its relative operational success compared with known alternative computational designs. This makes the assessment of intelligence relative to specified goals. Obviously, the best design will depend on which goal is selected. Different methods will perform best according to different definitions of success. "Goals" in this sense, include all of the different issues of costs and benefits relevant to alternative computational systems and decision consequences. For example, which kinds of errors are costly and which kinds are cheap (what, for example, is the cost of being afraid of a nonvenomous snake versus the cost of being unafraid of a venomous one)? What is the

cost (in time, metabolic energy, processing load, and so on) of one system of computation as opposed to another? Also, the best design will depend on the distribution of background conditions within which problem solving is to take place. Different designs will perform best in different problem-solving environments.

Natural problem solving tends to take place in complex environments with certain stable or statistically recurring features. To understand why a particular computational method will prove more effective in one environment than another, one needs to answer such questions as the following: What is always true in the task environment, what is statistically true, and what is never true? What do detectable cues predict about the undetectable features of the environment? What information is routinely available? How stable are the variable dimensions of the task environment? And so on. Moreover, the best design will depend on the ecological distribution of different problem types that the problem-solving system encounters. Because computational strategies ordinarily involve trade-offs, different methods will perform best against different composite problem populations. Thus, the answer to the question, Which design is most intelligent?, is not and cannot be invariant and universal. The intelligence of a design is always relative to the goal to be reached (or the total array of values and trade-off functions), to the background conditions that it operates in, to the total problem population to which it will be applied, and to other factors as well. We have called the well fittedness of computational designs to environments *ecological rationality* (Tooby & Cosmides, 1992a; see also Gigerenzer, Todd, & ABC Research Group, 1999).

On the other hand, this definition of dedicated intelligence differs from more traditional views in a series of ways. For example:

- *It privileges adaptive problems over other kinds of problems.* Adaptive problems are the enduring cause-and-effect relationships that select for some design features over others. If we are to understand in what way mechanisms in our minds were designed to be intelligent, we need to relate these designs to the structure of the problems they were constructed to solve. In contrast, the pursuit of nonadaptive outcomes by an organism is a by-product of computational machinery designed by natural selection for the realization of adaptive outcomes. A male robin red breast may not look particularly intelligent when it overcomes obstacles to attack a tuft of red features, nor does a human male when he spends time looking for pornographic pictures rather than courting actual women (Dawkins, 1982). But the computational systems that organize the behavior of the robin and the man such that they pursue these goals exhibit intelligence$_1$ nevertheless. These mechanisms lead to such odd outcomes because there are things in the world other than rival robins and living women that satisfy the input conditions for the monitoring devices employed by the computational systems that (respectively) regulate aggression in robins and courtship in humans (see Sperber, 1994, on the actual versus proper domain of an adaptation).
- *It is easily applied to organisms but does not apply as easily to human-made machines.* Because natural selection applies generally to anything capable of self-reproduction and mutation, this approach to intelligence can be used to recognize instances of intelligence$_1$ in any species. Because human-made artifacts are not themselves replicators, this definition cannot be directly applied to them. An analogue of this definition can be applied, if one is willing to specify a function for the machine. Alternatively, one could choose to look at artifacts as extensions of the human phenotype, as Dawkins (1982) does, which would then make their intelligence dependent on how well they served evolved goals. As Richard Dawkins has argued, machines are created to realize the goals of the organisms that designed them, and any intelligence exhibited by a machine was derived from the adaptations of the organisms that created it (Dawkins, 1982). In a similar vein, Dennett (1987) has argued that machines manifest "derived intentionality": a goal-directedness derived from the goals and intentions of the organism that made it, which manifests "original intentionality."
- *Because it privileges adaptive problems, it is difficult to apply the concept of dedicated intelligence to a system that executes complex behaviors to solve arbitrarily chosen problems.* Consider, for example, a person with autism who spends all his time memorizing the telephone book. Is this intelligent behavior or not? True, he is overcoming obstacles to achieve a goal, but it is an odd goal, unconnected to the solution of any ancestral adaptive problem, and it is pursued at the expense of nearly all other goals. This is the kind of situation for which the term *idiot savant* was coined: Such a person exhibits some features of intelligence but not others. On the other hand, if you discovered that this person was in fact a visitor from another planet, and that prior visitors had encrypted the coordinates of his home planet in the phone book, the same behavior would seem more intelligent, in part because returning home is an instance of an intelligible adaptive problem.

HOW IS DEDICATED INTELLIGENCE ACHIEVED?

All animals, including humans, are endowed with computational systems that manifest intelligence$_1$. Although this point is subject to a great deal of debate, we would argue that the human mind is very similar to the minds of other animal species. That is, it is bristling with a large number of specialized computational systems, each of which is well designed for solving a different adaptive problem. Functional specialization is one of the primary means by which computational systems achieve their problem-solving power, thereby manifesting intelligence$_1$.

Functional Specialization

Why should this be true? A basic engineering principle is that the same device is rarely capable of solving two different problems equally well. We have both screwdrivers and saws because each solves a particular problem better than the other. It would be futile to cut planks of wood with a screwdriver or to turn screws with a saw.

For exactly the same reason, natural selection has divided our body into organs such as the heart and the liver. Pumping blood throughout the body and detoxifying poisons are two very different problems. Consequently, your body has evolved a different machine for solving each of them. The design of the heart is specialized for pumping blood; the design of the liver is specialized for detoxifying poisons. Your liver can't function as a pump, and your heart cannot detoxify poisons.

The same principle applies to the mind. When carefully considered, it leads to the conclusion that the mind has many independent, evolved programs. One reason for this becomes clear if you put yourself in the position of a superhuman engineer. Imagine you are trying to design an organism like ourselves—one that has values and uses them to make choices. What would your organism be like if you gave it only one set of choice criteria?

Let's say your science project is to design a model human female, and you want her to be able to choose nutritious foods. Natural selection has engineered into humans an elaborate set of neural circuits organized to choose nutritious food on the basis of taste, smell, and digestive consequences. Knowing this, you decide to give your science project the same programs. But if this is the only set of choice criteria she has, what kind of *mate* would she end up choosing? A goat cheese pizza or a giant chocolate bar? Although superior to a bad date, they will not measure up as a parent to her children. To solve the adaptive problem of finding the right mate, her mental machinery would have to be guided by qualitatively different standards and values than when she is choosing the right food, or the right word, or the right path to get home.

We humans solve many different adaptive problems well. To accomplish these feats, *there must be at least as many independent evolved mental programs as there are adaptive domains in which the standards for successful behavior are qualitatively different.* We think that one can identify hundreds or perhaps even thousands of these domains, ranging from thermoregulation, parenting, and food choice to mate choice, friendship maintenance, language acquisition, romantic love, pollutant avoidance, predator defense, sexual rivalry, status attainment, projectile accuracy, and kin welfare. Since environments cannot provide organisms with definitions of problem-solving success, independent problem solvers must be built in to the brain for each incommensurate value domain. For this and many other reasons, the brain must be composed of a large collection of evolved circuits, with different circuits specialized for solving different problems. In this view, the brain is necessarily a diverse collection of dedicated computers networked together.

Functional specialization can take many forms. For choice behavior, knowledge of the appropriate criteria must somehow be embodied in the program, either as a database or implicitly, in the nature of the cues to which the procedures that cause attraction, repulsion or disinterest respond. But information about proximal goals is not the only kind of functional specialization that one sees in the mind. Biological machines are tailored to the structure of the environments in which they evolved, and information about the stably recurring properties of these ancestral worlds can be embodied in the very way their procedures work. For example, one function of vision is object recognition, and this is easier if the same object—e.g., a banana—appears to have the same color—yellow—from one situation to the next, regardless of changes in the wavelengths of the light illuminating it. This is called *color constancy*, and our visual system does it very well. Natural selection has created color constancy circuits that automatically compensate for the wild changes in illumination that occur on the surface of the earth as the sun traverses the sky and under variations in cloud cover and forest canopy (Shepard, 1992). As a result, that banana looks yellow to us at high noon and at sunset, even though, objectively speaking, it is swamped by red light at sunset, such that it is a source of far more red than yellow light. Natural—that is, ancestrally recurrent—changes in terrestrial illumination pose no problems for these circuits, because they are calibrated to them: Their procedures were shaped by them and embody knowledge about them. But these circuits cannot compensate for evolutionarily novel changes in illumination, such as the unearthly spectrum cast by the sodium vapor lights that illuminate many parking lots at night. The cars that we think of as red and green and blue all look a muddy brown when they are illuminated by these golden lights because our color constancy mechanisms were not shaped by, and embody no knowledge of, the spectral properties of sodium (Shepard, 1992).

Evolved Crib Sheets

This principle applies not just to perception but to all of our learning and reasoning circuits as well. In this view, many dedicated intelligences are equipped with design features that function as crib sheets. They come to a problem already "knowing" a great deal about it. This allows them to be far more intelligent than they otherwise would be if they embodied no equivalent to innate knowledge. For example, a newborn's brain has response systems that expect faces to be present in the environment; babies less than 10 minutes old turn their eyes and head in response to facelike patterns but not to scrambled versions of the same pattern (Johnson & Morton, 1991). Neural maturation brings other evolved circuits on line subsequently. [as the phrase doesn't add anything for those in the know, and is likely to be obscure to those 'not in the know', we'd rather leave it out]. Infants have strong assumptions, deriving from the evolutionary past, about how the world works and what kinds of things it contains, even at 2½ months (the point at which they can see well enough to be tested). They assume, for example, that the world will contain rigid objects that are continuous in space and time, and they have preferred ways of dividing the world into separate objects (Spelke, 1990). Indeed, an infant's mind is designed to *privilege* some hypotheses about what counts as an object over others. Ignoring shape, color, and texture (all of which they can see), they treat any surface that is cohesive, bounded, and that moves as a unit as a single object. Another privileged hypothesis is that solid objects are impenetrable (Baillargeon, 1986). So when one solid object appears to pass through another, these infants are surprised, just as you or I would be.

A baby with a completely open mind—one lacking any privileged hypotheses—would be undisturbed by such displays. Why shouldn't a toy train travel smoothly through a solid block of wood? If the superhuman engineer were to remove these privileged hypotheses from the baby's mind, the baby would be left without informative guidance in the world in which we actually live. By definition, a blank-slate system must entertain all possible hypotheses equally: that it was born into a world in which objects are like mercury droplets, no one has a face, and surfaces that move together are physically unconnected to each other. These are properties of imaginable universes but not of the one in which we evolved. There is nothing in our evolutionary past that would cause our brains to be organized in such a futile way.

So babies have dedicated intelligences built into them with strong commitments about the nature of the universe and niche they actually evolved in, instead of being prepared to deal with all worlds, whether they exist or not. In watching objects interact, babies less than a year old distinguish causal events from noncausal ones that have similar spatio-temporal properties (Leslie, 1988, 1994); they distinguish objects that move only when acted upon from ones that are capable of self-generated motion (making the inanimate/animate distinction) (Gergely, Nadasdy, Csibra, & Biro, 1995; Mandler & McDonough, in press; Premack & Premack, 1997), and they assume that the self-propelled movement of animate objects is caused by invisible internal states—goals and intentions (Baron-Cohen, 1995). Toddlers have a well-developed mind-reading system (i.e., a system for intuiting what is on others' minds), which uses eye direction and movement to infer what other people want, know, and believe. This system is domain-specific: It is designed only for understanding the behavior of animate beings. It is content-dependent: It is activated by stimuli that have properties ancestrally associated with animate beings, such as eyes or self-propelled motion (seeing a boulder rarely excites curiosity about its hopes, ambitions, or beliefs). And it is functionally-specialized: It is designed to compute beliefs, desires, and intentions, not color, trajectory, or weight. Indeed, the mind-reading system is so functionally specialized that it can be selectively impaired (i.e., impaired while other cognitive abilities are intact). This can be clearly seen in certain people with autism (Baron-Cohen, 1995; Leslie, 1987).

The Structure of a Dedicated Intelligence

The structure of a dedicated intelligence reflects the long-enduring structure of the adaptive problem it solves. Natural selection coordinates the structure of a recurrent adaptive problem (including the features of the environment in which it occurs) with the structure of an adaptive problem solver such that the interaction of the two produces the solution to the problem. If selection has created a well-engineered adaptation, then elements that are necessary to solve the problem but lacking from the world are supplied by the structure of the problem-solving device. Equally, that which is reliably supplied by the environment will tend to be left out of the device, because too much redundancy will be unnecessarily costly. So, strictly speaking, one should not look for the complete solution to the adaptive problem in the mechanism itself; the solution emerges from the complementary interaction of the mechanism and the world. For example, the visual system supplies exactly the information about the world (in the form of assumptions built into scene analysis) that the retina is incapable of supplying (Marr, 1982). Linguistic evidence available to the child supplies too few constraints to allow grammar acquisition to proceed, so the language acquisition device makes assumptions about grammar that are present in the structure of all known human languages (Pinker & Bloom, 1990). To understand the operation and organization of our dedicated intelligences, it is necessary to understand what regularities reliably permeated the structure of natural problem environments—the environment of evolutionary adaptedness, or EEA (Tooby & Cosmides, 1990, 1992b). Obviously then, the malfunctioning of our dedicated intelligences frequently comes about when a situation lacks cues and relationships that tended to be stably true in the past, and on which the intelligence relies for its successful operation. This is why one must talk about the ecological rationality of evolved computational devices; no intelligent architecture can operate properly outside of the context for which it was designed (Gigerenzer et al., in press; Tooby & Cosmides, 1992a).

Dedicated Intelligences Expand Our Abilities

In the past, many researchers have assumed that violations of the blank-slate assumption would limit intelligence. However, autism graphically illustrates what happens when an evolved intelligence is missing. A person with autism may have a normal IQ, be better than normal at embedded figures tasks (like *Where's Waldo?*), and be able to make sophisticated inferences about machines. Yet this same person cannot make simple inferences about other people's beliefs and desires. If a normal 3-year-old sees a character, Charlie, looking at one of four candies and is asked, "Which candy does Charlie want?", the child will point to the one Charlie's eyes are trained on. But a person with autism will answer randomly, even though he can tell you exactly which candy Charlie is looking at (Baron-Cohen, 1995). The person with autism can detect eye direction but, unlike you or me, he cannot use it to infer what someone wants. This shows that whatever the mental tool kit is that comes with having a normal IQ and normal abilities to reason about the physical world, it is not sufficient for reasoning about the mental world. Because the mind of a person with autism is missing a dedicated intelligence designed to make inferences about the mental world, he does not know that eye direction can indicate desire. Similarly, having an intact mind-reading system is insufficient for reasoning about the physical world: Adults with Williams syndrome are good at inferring other people's mental states, yet they are profoundly retarded and have difficulty learning even very simple spatial tasks (Tager-Flusberg, Boshart, & Baron-Cohen, 1998).

Domain-specialized inferential tools and knowledge bases are found not just in the learning systems of infants and toddlers, but in those of adults as well. For example, it is now well established (if not universally assented to) that the learning mechanisms that govern the acquisition of a language are different from those that govern the acquisition of food aversions, and both of these are different from the learning mechanisms that govern the acquisition of snake phobias. Each program has knowledge of its particular domain built into its structure, which allows it to perform its function far more efficiently than any blank-slate system could. The language acquisition device knows, for example, that the names of objects are nouns (Pinker, 1994). The snake phobia system knows what snakes look like, knows what fear looks like on other's faces, and has a procedure specialized for using fear on other's faces to change the intensity of fear you feel in the presence of snakes (Mineka & Cook, 1993; Ohman, Dimberg, & Ost, 1985). The food aversion system knows that nausea is usually caused by foods recently ingested, that it is more likely to be caused by novel foods than by familiar foods, and uses the contingency between food ingestion and nausea to regulate the subsequent attractiveness of food items (Garcia, 1990; Seligman, 1971). How did these systems get these specialized procedures and knowledge? Those mutations that, for example, built in the knowledge of what snakes looked like and what a fear-face looked like, increased the efficiency with which one learns which snakes should be avoided; hence, they were selected for.

The mind is not packed with specialized programs merely because they afford small differences in efficiency. Different problems *require* different dedicated intelligences. Knowledge about beliefs and desires, which allows one to infer the behavior of other people, will be misleading if it is applied to rocks and lakes. Knowing that concrete objects are nouns will not allow you to avoid venomous snakes. Two devices are better than one when the crib sheet that helps solve problems in one domain is misleading—or useless—in another. This is why many dedicated intelligences are designed to be activated in one domain and not others: To be useful, they must be activated only in those domains that match the assumptions they work under.

The more dedicated intelligences an architecture has, the more problems it can solve. A brain equipped with a multiplicity of specialized inference engines will be able to generate more successful types of problem-solving behavior than an architecture that is stripped of specializations. In this view, the flexibility and power often attributed to blank slates and content-independent algorithms is illusory. All else being equal, a content-rich system will be able to infer far more than a content-poor one.

Why Content-Rich Is Better Than Content-Poor

This view of the mind is radically at variance with the model of the mind that is the centerpiece of the standard social science model. Its advocates have attributed everything—from hopscotch to romance to rivalry—to the evenhanded operation of "learning", "intelligence", "reasoning", and "decision making." Regrettably, those simply remain names for mysterious hypothetical processes, not well-validated theories of how things actually happen computationally. To fill this gap, cognitive scientists proposed that the mind comes endowed with general-purpose computational circuits that are jacks-of-all-trades. Prime candidates were so-called *rational* algorithms: programs that implement formal methods for inductive and deductive reasoning, such as the laws of probability, mathematics, or formal logic. Others have proposed comprehensive pattern associator architectures that compute correlations or contingencies. These methods are inviting precisely because they are content free. Given the seemingly inexhaustible diversity of human action, it seemed reasonable to conclude that the mind be initially free of all content, so that variations in experience could drive the accumulation of the rich particularity so notable in the individual human mind.

What do we mean by a content-free program? Consider *modus ponens* and *modus tollens*, two domain-general rules of logic. Whenever "If P then Q" is true and P is true, *modus ponens* allows you to validly conclude that Q is also true. *Modus tollens* licenses a different inference: When "If P then Q" is true, but Q is false, it allows you to conclude that P is also false. These rules are content independent: They allow you (or an automaton, such as a computer or a neural circuit) to deduce true conclusions from true premises, no matter what is substituted in for P and Q. Let's say that P = you snooze and Q = you lose. If it is true that "If you snooze, you lose" then you can conclude that anyone who snoozed lost (*modus ponens*), and anyone who won didn't snooze (*modus tollens*). They will produce new knowledge whenever a true premise is combined with a true if-then statement—anything from "If it rains, the ground gets wet" to "If you can keep your head while all those around you are losing theirs, then you'll be a man, my son." Bayes's rule, a widely used equation for computing the probability that a hypothesis is true given data about that hypothesis, is also content independent. It can be applied equally to medical diagnosis, deciding whether Paul McCartney was dead before *Abbey Road* was recorded, playing Baccarat against James Bond, or any other subject matter.

Unfortunately, devices limited to executing Bayes's rule, modus ponens, and other "rational" procedures derived from mathematics or logic are computationally very weak compared with an evolved system of dedicated, content-specialized intelligences (Tooby & Cosmides, 1990, 1992b). The theories of rationality embodied by such traditional rational procedures, in order to be able to make valid inferences for all possible contents in all possible domains, have no built-in assumptions about the long-term ecological structure of the world or the problem domain (Gigerenzer et al., in press). They can be applied to a wide variety of domains, however, only because they lack any information that would be helpful in one domain but not in another. Having no evolved problem spaces or specialized procedures tailored to a domain, there is little they can deduce about it; having no privileged hypotheses, there is little they can induce before their operation is hijacked by combinatorial explosion—the cost of considering, searching, or processing all of the combinatorial possibilities. These jacks of all trades are, necessarily, masters of none. They achieve generality only at the price of broad ineptitude. Domain-specific algorithms do not need to make the same trade-off: Each can be master of a different domain. The difference between domain-specific methods and domain-independent ones is akin to the difference between experts and novices: Experts can solve problems faster and more efficiently than novices because they already know a lot about the problem domain.

Dedicated intelligences—such as the ones that govern how we reason and learn about faces, objects, language, snakes, mind reading, nausea, and so on—have the following five properties (Pinker, 1994):

1. they are complexly structured for solving a specific type of adaptive problem;
2. they reliably develop in all normal human beings;
3. they develop without any conscious effort and in the absence of any formal instruction;
4. they are applied without any conscious awareness of their underlying logic; and
5. they are distinct from whatever more general abilities to process information or behave intelligently that may exist.

In short, they have all the hallmarks of what scholars would once have called an instinct (Pinker, 1994). To reconnect cognitive science with evolutionary biology, these functionally specialized, content-rich intelligences can be considered reasoning instincts and learning instincts. They make certain kinds of inferences just as easy, effortless, and natural to humans as spinning a web is to a spider or dead reckoning is to a desert ant. In short, instincts manifest intelligence$_1$: They are well designed for solving adaptive computational problems.

For most of this century, the consensus has been that even if other animals are ruled by "instinct," humans have lost their instincts and had them replaced with "reason," "intelligence," or "learning." This evolutionary erasure and substitution is the explanation for why humans are more flexibly intelligent than other animals. William James (1892), however, argued against this commonsense view. He maintained that human behavior is more flexibly intelligent than that of other animals because we have more instincts than they do, not fewer. If instincts are like tools in a toolbox, then the larger the number that the mind is endowed with, the more abilities it has. James' view fits presciently with work in modern computer science, in which each additional subroutine expands the computer's ability to solve problems.

There is no reason to think that instincts are what we have in common with other species, whereas what is uniquely human is noninstinctual. Not only are instincts or dedicated intelligences often specific to each species, but many of our instincts give rise to abilities that are unique to humans, such as language. As Darwin put it, humans manifest language because we evolved "an instinctive tendencies to acquire an art" (see Pinker, 1994, p. 20).

Finally, we think that having a brain that is well endowed with computational systems that manifest intelligence$_1$ is a precondition for the evolution of intelligence$_2$, improvisional intelligence. To pick one necessary contribution, dedicated intelligences prevent combinatorial explosion and create a context in which design features that increase flexibility—a dangerous addition—can continue to have adaptive functional consequences.

BEYOND DEDICATED INTELLIGENCE: THE HOMINID ENTRY INTO THE COGNITIVE NICHE

When contextualized within the extraordinary diversity of the living world, humans stand out, exhibiting a remarkable array of strange and unprecedented behaviors—from super tankers to ice skating to sculpture—that are not found in other species. What is at the core of these differences? Arguably, one central and distinguishing innovation in human evolution has been the dramatic increase in the use of contingent information for the regulation of improvised behavior that is successfully tailored to local conditions—an adaptive mode that has been labeled the *cognitive niche* (Tooby & DeVore, 1987). If you contrast, for example, the food acquisition practices of a bison with that of a !Kung San hunter, you will immediately note a marked difference. For the bison, grasslands are undoubtedly a rich tapestry of differentiated food patches and cues; nevertheless, the bison's decisions are made for it by dedicated intelligences designed for grass and forage identification and evaluation—adaptations that are universal to the species and that operate with relative uniformity across the species range. In contrast, the !Kung hunter uses, among many other nonspecies-typical means and methods, arrows that are tipped with a poison found on only one local species of chrysomelid beetle, toxic only during the larval stage (Lee, 1993).

This method of food acquisition is not a species-typical adaptation: Not all humans use arrows, poison their arrows, have access to a beetle species from which poison can be derived, or even hunt. Nor are any of the component relationships—between beetle larva and poison, between arrows and poison, or even between arrows and hunting—stable from a phylogenetic perspective. Each relationship on which this practice is based is a transient and local condition, and these contingent facts are being combined to improvise a behavioral routine that achieves an adaptive outcome: obtaining meat. Whatever the neural adaptations that underlie this behavior, they were not designed specifically for beetles and arrows but exploit these local, contingent facts as part of a computational structure that treats them as instances of a more general class (e.g., living things, projectiles, prey).

Most species are locked in coevolutionary, antagonistic relationships with prey, rivals, parasites and predators, in which move and countermove take place slowly, over evolutionary time. Improvisation puts humans at a great advantage: Instead of being constrained to innovate only in phylogenetic time, they engage in ontogenetic ambushes against their antagonists—innovations that are too rapid with respect to evolutionary time for their antagonists to evolve

defenses by natural selection. Armed with this advantage, hominids have exploded into new habitats, developed an astonishing diversity of subsistence and resource extraction methods, caused the extinctions of many prey species in whatever environments they have penetrated, and generated an array of social systems, artifacts, and representational systems far more extensive than that found in any other single species.

This contrast—between transient, local, contingent facts and relationships that hold over the species range—is at the heart of what makes humans so different. To evolve, species-typical behavioral rules must correspond to features of the species' ancestral world that were both globally true (i.e., that held statistically across a preponderance of the species' range) and stably true (i.e., that remained in effect over enough generations that they selected for adaptations in the species). These constraints narrowly limit the kinds of information that such adaptations can be designed to use. The set of properties that had a predictable relationship to features of the species' world that held widely in space and time is a very restricted one. In contrast, for situation-specific, appropriately tailored improvisation, the organism only needs information to be applicable, or "true," temporarily, locally, or contingently. If information only needs to be true temporarily, locally, and situationally to be useful, then a vastly enlarged universe of context-dependent information becomes potentially available to be employed in the successful regulation of behavior. This tremendously enlarged universe of information can be used to fuel the identification of an immensely more varied set of advantageous behaviors than other species employ, giving human life its distinctive complexity, variety, and relative success. Hominids entered the cognitive niche, with all its attendant benefits and dangers, by evolving a new suite of cognitive adaptations that are evolutionarily designed to exploit this broadened universe of information, as well as the older universe of species-extensive true relationships.

The hominid occupation of the cognitive niche is characterized by a constellation of interrelated behaviors that depend on intensive information manipulation and that are supported by a series of novel or greatly elaborated cognitive adaptations or dedicated intelligences. This zoologically unique constellation of behaviors includes locally improvised subsistence practices; extensive context-sensitive manipulation of the physical and social environment; "culture," defined as the serial reconstruction and adoption of representations and regulatory variables found in others' minds through inferential specializations evolved for the task; language as a system for dramatically lowering the cost of communicating propositional information; tool use adapted to a diverse range of local problems; context-specific skill acquisition; multi-individual coordinated action; and other information-intensive and information-dependent activities (Tooby & Cosmides, 1992b).

Although some have argued that social competition was the sole driving force behind the evolution of human intelligence (as in the Machiavellian hypothesis; Humphrey, 1984; Whiten & Byrne, 1997), we do not think this is a sufficient explanation for what is distinctive about human intelligence (for an alternative, see Tooby & DeVore, 1987). We certainly do believe that humans have evolved dedicated intelligences specialized for social life and social cognition (e.g., Cosmides, 1989; Cosmides & Tooby, 1989, 1992), but what is truly distinctive about human life encompasses far more than the social. For example, the causal intelligence expressed in hunter-gatherer subsistence practices appears to be as divergent from other species as human social intelligence. So, improvisational intelligence is not simply dedicated social intelligence—something we also know from the fact that individuals with autism can lose social intelligence while maintaining high levels of causal intelligence.

WHAT IS IMPROVISATIONAL INTELLIGENCE?

Earlier, we defined intelligence$_2$ as intelligence$_1$ plus enhancements. More specifically, we said that a system is intelligent$_2$ to the extent that it is well designed for solving adaptive computational problems *and has components designed to exploit transient local conditions to achieve adaptive outcomes*. Whether in social interactions, hunting, toolmaking, programming, poetry, legal argumentation, athletics, or anything else, people recognize the presence of a distinctively human kind of intelligence when people reach goals more effectively through the tailoring of their conduct to take into account the distinctive features of the situation they are in. The rigid application of rules, regardless of whether they seem appropriate to the circumstances, and regardless of their success at reaching goals, strikes humans of whatever culture as diagnostic of a lack of intelligence.

Dedicated intelligence seems directly related to adaptive problems (nutrition, relationships, perception), while it is less obvious that the same is true for improvisational intelligence. The reason for this is that, in improvising to reach an adaptive outcome (e.g., Zorro defeating his enemies), one may need to pursue any of an endless array of intermediate goal states without intrinsic reward characteristics (e.g., Zorro playing the fool to keep his identity hidden). Hence, a system that evolves toward improvisional intelligence will produce, as a by-product, a system that can also compute how to pursue a large body of seemingly arbitrary goal states that are not necessarily adaptive. This is why improvisional intelligence appears to resemble the traditional concept of a general-purpose intelligence, despite the differences in conceptions of the machinery that achieves this outcome. This also makes it obvious why two problems that confront the evolution of improvisational intelligence are (i) the need to keep its use coupled to adaptive goals, and (ii) producing inferences that are correct (or, at least, useful) sufficiently often to pay for its cost.

The benefits of successful improvisation are clear: The ability to realize goals through exploiting the unique opportunities that are inherent in a singular local situation yields an advantage over a system that is limited to applying only those solutions that work across a more general class of situation. What 10 years of ordinary battle on the plains of Troy could not accomplish, one Trojan horse could. The improvisational exploitation of unique opportunities also fits our folk intuitions about what counts as intelligence. As members of the human species, instances of intelligence excite our admiration precisely to the extent that the behavior (or insight) involved is novel, and not the result of the "mindless" application of fixed rules. Indeed, it would seem that every organism would be benefitted by having a faculty that caused it to perform behaviors fitted to each individual situation. But: If it is generally useful, why haven't many other species evolved this form of intelligence (Tooby & DeVore, 1987)? Indeed, how is this form of intelligence computationally and evolutionarily possible at all?

To see why the existence of this form of intelligence is puzzling, let us first consider what is meant by conditions that are transient and local and the difficulty of building adaptations to the transient.

For an allele to spread to fixation throughout a species, it is not enough for the design feature it builds to confer an advantage in a single lifetime or a single locale. The incorporation of a trait into a species' design by selection is a large-scale, cumulative process, involving the summation of events that take place across the entire species' range and across a large number of generations. For selection to propel an allele consistently upwards, the relevant relationships between the environment, the organism, and the adaptive benefit must be stable—they must persist across many generations. For this reason, the functional designs of species-typical computational adaptations should, in general, both reflect and exploit conditions that hold true over long periods of time and over most or all of the species range. For example, eye direction statistically signals knowledge acquisition in organisms with eyes, and so monitoring eye direction is informative for making inferences about what another organism knows (i.e., seeing is knowing; Baron-Cohen, 1995). The mechanisms that make these inferences are components of a system that achieves adaptive outcomes by exploiting conditions that are stable with respect to the phylogenetic history of our species, even though these conditions are experienced as transient and local by individual human beings.

This stability can, of course, be of a statistical nature. Undoubtedly there are many cases in which a predator fails to recognize something it is looking at (otherwise camouflage would not have evolved in so many prey species). But the correlation between eye direction and object recognition can be weak, as long as it is positive; all that is necessary for selection to favor an eye direction detector is that using eye direction to infer knowledge confer a reproductive advantage—however slight—over not using it. Reliably occurring variations also count as stable relationships that selection can exploit. As we discussed, the human color constancy system is designed to compensate for wide variations in terrestrial illumination. True, the spectral properties of the light you experience are transient over the course of a day, and differ from location to location. But they are evolutionarily recurrent variations. They are not transient and local in the sense intended in the definition of intelligence$_2$. And the color constancy system exhibits intelligence$_1$ but not intelligence$_2$. It produces color constancy when illuminant conditions fall within the envelope of variations that were stably present during the evolution of this system (intelligence$_1$), but it cannot exploit conditions that are evolutionarily transient and local, such as the spectral properties of the sodium vapor lamp, to produce the adaptive outcome of color constancy. Thus, *transient* and *local* are here defined with respect to the history of a species, not the history of an individual.

THE ENIGMA OF IMPROVISIONAL INTELLIGENCE

The costs and difficulties of the cognitive niche are so stringent that only one lineage in 4 billion years has wandered into the preconditions that favored the evolution of this form of intelligence. Natural computational systems that begin to relax their functional specificity run into, and are inescapably shaped by, savagely intense selection pressures. One of the greatest problems faced by natural computational systems is combinatorial explosion (for discussion, see Cosmides & Tooby, 1987; Tooby & Cosmides, 1992b). Combinatorial explosion is the term for the fact that alternatives multiply with devastating rapidity in computational systems, and the less constrained the representational and procedural possibilities are, the faster this process mushrooms, choking computation with too many possibilities to search among or too many processing steps to perform. Every marginal increase in the generality of a system exponentially increases the computational cost, greatly limiting the types of architectures that can evolve, and favoring, for example, the evolution of modules only in domains in which an economical set of procedures can generate a sufficiently large and valuable set of outputs. This means that domain specificity—and dedicated intelligences—will be the rule rather than the exception in natural computational systems. And while it answers the question of why a broad, general form of intelligence is so extraordinarily rare among animal species, it deepens the question of how it could be possible at all.

Elsewhere, we have written at length about the trade-offs between problem-solving power and specialization: general-purpose problem-solving architectures are very weak but broad in application, whereas special-purpose problem-solving designs are very efficient and inferentially powerful but limited in their domain of application (Cosmides & Tooby, 1987; Tooby & Cosmides, 1992b). Thus, on first inspection, there appear to be only two biologically possible choices for evolved minds: either general ineptitude or narrow competences. This choice appears to rule out general intelligence. Yet, hominids did manage to evolve an architecture that allowed them to enter the cognitive niche, exploiting conditions that, from a phylogenetic perspective, are transient and local, to achieve adaptive outcomes. What is the way out of this puzzle?

We cannot simply return to the traditional view. The traditional argument that because human intelligence appears unprecedentedly broad in application, the human cognitive architecture's core problem-solving engines must themselves be general purpose, cannot be reconciled with what is now known about the complexity of natural problems and the shortcomings of such architectures.

Nor have we yet confronted the core of the problem. From evolutionary and computational perspectives, it is far from clear how local improvisation could evolve, operate, or even be a nonmagical, genuine cognitive possibility. The central evolutionary enigma behind improvisional intelligence can be stated as follows: A computational system, by its nature, can only apply rules or procedures to problems and must do so based on its categorization of individual problems into more general classes (i.e., there must be a causal process whereby appropriate procedures are activated in a given situation).[1] Adaptations, by their nature, can only see individual events in the life of the organism as instances of the large-scale evolutionarily recurrent categories of events that built them (Tooby & Cosmides, 1990). Therefore, if computational systems can only respond to situations as members of classes to which computational rules apply, and if evolution only builds computational adaptations that see individual situations as members of large-scale, evolutionarily recurrent classes of events, how can there be a brain whose principles of operation commonly lead it to improvise behaviors that exploit the *distinctive* features of a situation? By the nature of how selection works, how could species-typical computational rules evolve that allow situation-specific improvisation at all, much less at a sufficiently low cost?

These are all difficult problems, and we suspect that no one presently has a full account of how improvisional intelligence could evolve and what subcomponents it requires for its operation. However, we think there are some tentative answers that look promising.

To start, there is an alternative to domain-general ineptitude or narrow intelligence. Cognitive specializations, each narrow in their domain of application,

[1]By rules or procedures, we only mean the information-processing principles of the computational system, without distinguishing subfeatural or parallel architectures from others.

can be bundled together in a way that widens the range of inputs or domains that can be successfully handled. This avoids the weakness of an architecture that consists of content-independent procedures, while avoiding the narrowness of a single domain-specific inference engine. It gets the benefits of specialized problem-solving power but progressively widens the scope of the problems that can be solved with each additional specialization that is added.

Moreover, such an architecture can be further improved; compatible content-independent engines can be embedded within this basic design because their defects when operating in isolation can be offset by implanting them in a guiding matrix of specializations (e.g., Cosmides & Tooby, 1996b; Gigerenzer et al., 1999). For example, the specializations provide the input content and the problem spaces, choking off combinatorial explosion, and provide a large repertoire of efficient specialized inference rules to augment the general inference rules. Of course, other architectural features are required to solve the problems raised by the interactions of these heterogeneous systems, as discussed later in this chapter (Tooby & Cosmides, 1990, 1992a, 1992b). This seems to us to be a necessary if partial solution to the question of how human intelligence can be not only broad in its range of application but also sufficiently powerful when applied (Sperber, 1996; Tooby & Cosmides, 1990, 1992b).

Second, a promising answer to the question of how evolved mechanisms, which are built only by species-wide regularities, can evolve to represent the distinctive or unique aspects of individual situations might be as follows: All situations are decomposed according to evolved interpretive rules that do see its elements only as instances of evolutionarily recurrent categories. (There seems to be no other possibility.) However, any given situation can be represented as unique in its particular combination of evolutionarily recurrent elements. The elements are computationally meaningful as instances of evolved categories, which allows evolved problem-solving rules to be applied to them. Indeed, the more evolved categorization systems that intersect on the same situation, the more situation interpretations are possible, and the more alternative manipulations can be considered and sifted according to evaluation systems that recognize valuable outcomes. (So, for example, we have the choice of viewing a man as a physical object, as an animal, as an agent with mental states, as a son, a potential sex partner, a shape, a member of a species, and so on.) Thus, the behavioral course decided upon might be uniquely tailored to the local situation, not because the elements are interpreted as novel but because the configuration taken as a whole is a novel combination of familiar elements. On this view, improvisional intelligence would benefit from a familiarity with the elements involved in unique situations, and should be stalled when genuinely new elements appear (which seems to be accurate).

Third, improvisional intelligence does not appear to be an autonomous ability, disconnected from the rest of the architecture and not relying on any other computational or informational resource. On the contrary. Not only does it depend on a base of dedicated intelligences but it also must be supplied with a *dense accumulation of information relevant to the situation being faced.* This is why we emphasized that the hominid entry into the cognitive niche depended on the huge increase in the use of *contingent* information for the regulation of improvised behavior that is successfully tailored to local conditions. The intensive use of information that is only temporarily or locally true creates what we have called the *scope problem.* Hence, we think another aspect to improvisional intelligence is a series of computational adaptations—what we have called *scope syntax*—to solve the problems introduced by the exploitation of contingent information. We think that any system that humans would recognize as having intelligence$_2$ will have a scope syntax.

WHAT IS THE SCOPE PROBLEM?

When hominids evolved or elaborated adaptations that could use information based on relationships that were only "true" temporarily, locally, or contingently, this opened up a new and far larger world of potential information than was available previously. Context-dependent information could now be used to guide behavior to a far greater extent than had been possible before. This advance, however, was purchased at a cost: *The exploitation of this exploding universe* of potentially representable information creates a vastly expanded risk of possible misapplications. This is because information that is useful within a narrow arena of conditions can be false, misleading, or harmful outside of the scope of those conditions[2]. Imagine, for example, that the following piece of contingent information is true: "The mushrooms [*here*] are edible." This is useful if you are collecting food here. But if you are collecting food elsewhere, this same information could kill you: The mushrooms *here* might be edible, but the mushrooms 3 miles away may be poisonous. To be useful, there needs to be a way of representing the scope within which the information about mushrooms being edible is true; *here* is a scope marker. (We represented this scope marker with a word, but to be useful in guiding an individual's behavior, it only needs to take the form of a conceptual tag attached to the information.) Or consider a different kind of contingent information, this time pertaining to someone's beliefs: "[*Bo believes that*] his arrows are back at camp." You can use this piece of contingent information to predict where Bo will go to look for his arrows, even if, in re-

[2]Indeed, the world outside of the local conditions may be commonly encountered, and depending on how narrow the envelope of conditions within which the information is true, scope-violating conditions are likely to be far more common than the valid conditions.

ality, someone stole them and hid them near the stream. If you want to retrieve the arrows yourself, you do not go to the camp. *Bo believes that* acts as a scope marker, which represents the boundaries within which the information about arrows can be usefully applied. This scope marker makes a very limited guarantee: it tells you the information will be useful for predicting Bo's behavior, but it can't promise that it will be useful for other purposes, such as fetching arrows. (Leslie [1987] calls data formats with slots for an agent [*Bo*], an attitude [*believes*] and a proposition [*his arrows are back at the camp*] a *meta*representation, because it is a representation that is about another representation—in this case, one that is in Bo's head.)

The cognitive niche depends on a computational strategy in which information is used even though the information is only applicable temporarily or locally. But this computational strategy can be successful only if the boundaries within which each representation remains useful are specified. Are the beetle larvae that are used to poison arrows toxic at all times of the year? Once harvested and applied, how long does the poisoned arrow tip remain poisonous? If it is poisonous to humans, gazelles, and duikers, is it also poisonous to lions, cape buffalo, and ostriches? If these relationships are true here, are they true on foraging territories on the other side of the Okavango? If the first several statements from my father in answer to these questions turned out to be true, will the remainder be true also? Moreover, because these aspects of the world are (by definition) transient and local, their boundaries must be continually monitored and reestablished.

Information only gives an advantage when it is relied on inside the envelope of conditions within which it is applicable. Hence, when considering the evolution of adaptations to use information, the costs of overextension and misapplication have to be factored in, as well as the costs and nature of the defenses against such misapplication. Expanding the body of information used to make decisions is harmful or dangerous *if* the architecture does not and cannot detect and keep track of which information is applicable, where it is applicable, and how the boundaries of applicability shift.

Moreover, the problem is not simply that information that is usefully descriptive only within a limited envelope of conditions will (by definition) be false or harmful outside of the scope of those conditions. The scope problem is aggravated by the fact that information is integrated and transformed through inferences. Information is useful to the extent it can be inferentially applied to derive conclusions that can then be used to regulate behavior. Inferences routinely combine multiple inputs through a procedure to produce new information, and the value of the resulting inferences depends sensitively on the accuracy of the information that is fed into them. For example, the truth of the conclusion that it will be better to move to an area where there is more game is dependent on the proposition that there is more game in the new location and on the implicit or explicit assumption that the necessary poisons for hunting can be obtained there as well.

Not only does inference combinatorially propagate errors present in the source inputs, but the resulting outputs are then available to be fed in as erroneous inputs into other inferences, multiplying the errors in successive chains and spreading waves. For example, if one wrong entry is made in a running total, all subsequent totals—and the decisions based on them—become wrong. This process has the potential to corrupt any downstream data set interacted with, in a spreading network of compounding error. The more the human cognitive architecture is networked together by systems of intelligent inference, and the more it is enhanced by the ability to integrate information from many sources,[3] the greater the risk is that valid existing information sets will be transformed into unreconstructable tangles of error and confusion. In short, the heavily inference-dependent nature of human behavior regulation is gravely threatened by erroneous, unreliable, obsolete, out-of-context, deceptive, or scope-violating representations.

Thus, it is not just the great increase in the use of contingent information that is important in understanding the human entry into the cognitive niche but the equally great increase in the permitted interaction among representations and representational systems. This increase is a double-edged sword: It offers great benefits in allowing many new inferences to be made, but it also aggravates the problem of data corruption—what scope syntax is designed to cope with. This increase in permissible interactions requires adaptations for translating information across mechanism boundaries and into common formats that make this interaction possible. The breadth of inferential interaction is important in understanding the distinctive aspects of the cognitive niche. Many representations in the human mind are not limited in their scope of application. They can be acted on by inference procedures that evolved to process information from other domains (as when inference procedures that evolved for making stone tools are applied to bone, a material of animal origin; Mithen, 1996), and they are allowed to inferentially interact with each other to a zoologically unprecedented degree (as when one's knowledge of bison anatomy and behavior affects how one fashions a tool for hunting them; Mithen, 1996). This is a pivotal element making such an architecture advantageous: Information can be made far more useful, if different items can be integrated into the same inferential structure, to produce new derivations. This phenomenon has been given various names—conceptual blending (Turner, 1996), conceptual integration (Sperber, 1994), domain sharing, or cognitive fluidity (Mithen, 1996). But

[3] i.e., to be de-encapsulated.

so far it is easier to point to examples of it than to provide a causal account of the machinery that produces it (for an interesting possibility, see Sperber, 1994).

In any case, the evolution of intelligence will depend critically on the economics of information management (see, e.g., Boyd & Richerson, 1985) and on the tools for handling information—that is, the nature of the adaptations that evolve to handle these problems. The net benefit of evolving to use certain classes of information will depend on the cost of its acquisition, the utility of the information when used, the damage of acting on the information mistakenly outside its area of applicability, and the cost of its management and maintenance. Because humans are the only species that has evolved this kind of intelligence, humans must be equipped with adaptations that evolved to solve the problems that are special to this form of intelligence.

HOW INTELLIGENCE$_2$ IS ACHIEVED

Scope Syntax, Truth, and Naïve Realism

For these reasons, issues involving not only the accuracy but also the scope of applicability of the information that the individual human acquires and represents became paramount in the design and evolution of the human cognitive architecture. We believe that there are a large number of design innovations that have evolved to solve the specialized programming problems posed by using local and contingent information, including a specialized scope syntax, metarepresentational adaptations, and decoupling systems. Indeed, we think that the human cognitive architecture is full of interlocking design features whose function is to solve problems of scope and accuracy. Examples include truth-value tags, source tags (self versus other; vision versus memory, etc.), scope tags, time and place tags, reference tags, credal values, operators embodying propositional attitudes, content-based routing of information to targeted inference engines, dissociations, systems of information encapsulation and interaction, independent representational formats for different ontologies, and the architecture and differential volatility of different memory systems.

Large amounts of knowledge are embodied in intelligent$_1$, domain-specific inference systems, but these systems were designed to be triggered by stimuli in the world. This knowledge could be unlocked and used for many purposes, however, if a way could be found to activate these systems in the absence of the triggering stimuli—that is, if the inference system could be activated by imagining a stimulus situation that is not actually occurring: a counterfactual. For example, by imagining a situation in which I left a knife near the counter's edge while the baby is toddling about the house, useful inferences about space, rigid object mechanics, biomechanics, intuitive biology, and intuitive psychology are unlocked, and a scenario unfolds before the mind's eye: "At counter's edge, the knife is within the baby's reach; she [will] see it, reach for it, and hurt herself" (*will* is a scope marker).

Given that our perceptions of the world are themselves mental representations, altering the architecture such that it can generate representations with the appropriate triggering features might not be too difficult to engineer—especially if the eliciting circumstance is itself a visual, tactile, kinesthetic, or proprioceptive percept, such as my seeing my hand about to put the knife on the counter (Tooby & Cosmides, 1990a). But for reasoning from this counterfactual situation to be useful, something else is needed. The premise—"the knife is at the counter's edge"—cannot be stored as something that has actually happened (if it has not), and the conclusion—"the baby hurt herself"—must be tagged as something that *could* happen but that *has not* happened.

In other words, one critical feature of a system capable of suppositional reasoning is the capacity to carry out inferential operations on sets of inferred representations that incorporate suppositions or propositions of conditionally unevaluated truth value, *while keeping their computational products isolated from other knowledge stores* (i.e., decoupled from them) until the truth or utility of the suppositions is decided, and the outputs are either integrated or discarded. This capacity is essential to planning, interpreting communications, employing the information communication brings, evaluating others' claims, mind reading, pretense, detecting or perpetrating deception, using inference to triangulate information about past or hidden causal relations, and much else that makes the human mind so distinctive. In what follows, we will try to sketch out some of the basic elements of a scope syntax designed to defuse problems intrinsic to the human mode of intelligence. By a scope syntax, we mean a system of procedures, operators, relationships, and data-handling formats that regulate the migration of information among subcomponents of the human cognitive architecture (for a fuller treatment, see Cosmides & Tooby, in press; also Leslie, 1987; Sperber, 1985).

To clarify what we mean, consider a simple cognitive system that we suspect is the ancestral condition for all animal minds and the default condition for the human mind as well: naïve realism. For the naïve realist, the world as it is mentally represented is taken for the world as it really is, and no distinction is drawn between the two. Indeed, only a subset of possible architectures are even capable of representing this distinction, and in the origin and initial evolution of representational systems, such a distinction would be functionless. From our external perspective, we can say of such basic architectures that all information found inside the system is assumed to be true, or is treated as true. However, from the point of view of the architecture itself, that would not be correct, for it would imply that the system is capable of drawing the distinction between true and false, and is categorizing the information as true. Instead, mechanisms in the architecture simply use the information found inside the system to regulate behavior and to carry out further computations. Whatever information is present in the system simply is "reality" for the architecture. Instead of tagging information as true or false—which seems so obvious to us—such basic architectures would not be designed to store false information. When new information is produced that renders old information obsolete, the old information is updated, overwritten, forgotten, or discarded. None of these operations require the tagging of information as true or false. They only involve the rule-governed replacement of some data by other data, just like overwriting a memory register in a personal computer does not require the data previously in that register be categorized as false. For most of the behavior-regulatory operations that representational systems evolved to orchestrate, there would be no point to storing false information, or information tagged as false. For this reason, there is no need in such an architecture to be able to represent that some information is true; its presence, or the decision to store it or remember it, is the cue to its reliability. In such a design, true equals accessible.

With this as background, and leaving aside the many controversies in epistemology over how to conceptualize what truth "really" is, we can define what we will call *architectural truth:* Information is treated by an architecture as true when it is allowed to migrate (or be reproduced) in an unrestricted or scope-free fashion throughout an architecture and is allowed to interact with any other data in the system that it is capable of interacting with. All data in semantic memory, for example, is architecturally true. The simplest and most economical way to engineer data use is for "true" information to be unmarked, and for unmarked information to be given whatever freedom of movement is possible by the computational architecture. Indeed, any system that acquires, stores, and uses information is a design of this kind. The alternative design, in which each piece of information intended to be used must be paired with another piece of information indicating that the first piece is true, seems unnecessarily costly and cumbersome. Because the true-is-unmarked system is the natural way for an evolved computational system to originate, and because there are many reasons to maintain this system for most uses, we might expect that this is also the reason why humans, and undoubtedly other organisms, are naïve realists. Naïve realism seems to be the likely starting point phylogenetically and ontogenetically, as well as the default mode for most systems, even in adulthood.

The next step, necessary only for some uses, is to have representations embedded within other data structures: metarepresentations (in a relaxed rather than narrow sense). For example, a cognitive architecture might contain the following structure: *The statement that "astrology is a science" is true*. This particular data structure includes a proposition (or data element) and an evaluation of

the truth of the proposition (or data element).[4] However, such structures need not be limited to describing single propositions. Although it is common, in talking about metarepresentations and propositional attitudes, to depict a single representation embedded in an encompassing proposition, a single proposition is only a limiting case. A set of propositions or any other kind of data element can be bundled into a single unit that is taken, as a data packet, as an argument by a scope operator to form a metarepresentation. For example, the metarepresentation *Every sentence in this chapter is false* describes the truth value of a set of propositions as easily as *The first sentence in this chapter is false* describes the truth value of a single proposition. Indeed, sometimes integrated sets of propositions governed by a superordinate scope operator might become so elaborated, and relatively independent from other data structures, that they might conveniently be called *worlds*. We think large amounts of human knowledge inside individuals exists inside data structures of this kind.

A sketch of the kind of cognitive architecture and operators we have in mind begins with a primary workspace that operates in a way that is similar, in some respects, to natural deduction systems (see Gentzen, 1935/1969; Rips, 1994; Cosmides & Tooby, 1996a), although it may include axiom-like elements and many other differences as well. Its general features are familiar: There is a workspace containing active data elements, and procedures or operators act on the data structures, transforming them into new data structures. Data structures are maintained in the workspace until they are overwritten, or if not used or primed after a given period of time, they fade and are discarded. Products may be permanently stored in appropriate subsystems if they meet various criteria indicating they merit long-term storage or warrant being treated as architecturally true. Otherwise, the contents and intermediate work products of the workspace are volatile and are purged, which is one adaptation for protecting the integrity of the reliable data stores elsewhere in the architecture (e.g., the fact that dreams are volatile is probably a design feature to avoid corruption of memory stores; Symons, 1993). Data structures may be introduced from perception, memory, supposition, or from various other system components and modules. Some of the procedures and tags available in the workspace correspond to familiar logical operators and elements, such as variable binding, instantiation, if introduction and if elimination, the recognition and tagging of contradictions, *modus ponens*, and so on. Some of the procedures are ecologically rational (Tooby & Cosmides, 1992a; Cosmides & Tooby, 1996b); that is, they correspond to licensed transformations in various adaptive logics (which may diverge

[4]There is no need, in particular, for the data structure to be a sentencelike or quasi-linguistic proposition. For most purposes, throughout this paper, when we use the term *proposition* we are not committing ourselves to quasi-linguistic data structures—we will simply be using it as a convenient short-hand term for a data element of some kind.

substantially from licensed inferences in the content-independent formal logics developed so far by logicians). Indeed, many procedures consist of routing data structures through adaptive specializations such as cheater detection or hazard management algorithms (Cosmides, 1989; Cosmides & Tooby, 1997), with outputs placed back into the workspace—a process that resembles either calling subroutines or applying logical transformations, depending on one's taste in formalisms.[5] Deliberative reasoning is carried out in this workspace, while many other types of inference are carried out automatically as part of the heterogeneous array of intelligent$_1$ specializations available in the architecture. Some areas of this workspace are usually part of conscious awareness, and most are consciously accessible.

Scope Representations

The data sets in this system exist in structured, hierarchical[6] relations, which we will represent as indented levels. Data elements in the left-most position are in what might be thought of as the ground state, which means they are licensed to migrate anywhere in the architecture they can be represented. Through inference procedures, they can mate promiscuously with any other ground-state data elements, producing conclusions that are their inferential offspring. Usually, ground-state elements are permitted to interact with subordinate levels as well. In other words, they are architecturally true, or scope free. Other elements are subordinated under ground state elements through scope operators. Therefore, we might represent an architecturally true statement in the left-most position:

(1) *Anthropology is a science.*

When in the left-most position, the statement is unmarked by the architecture. As such, it is free to be stored or to be introduced into any other nondecoupled process in the architecture. A subordinated statement may be scope limited, such as the following:

[5]Various operators and features of the workspace provide the intuitions that logicians have elaborated into various formal logics—the elaboration taking place through the addition of various elements not found in the workspace, the attempt to simultaneously impose self-consistency and conformity to intuition, and the removal of many content-specific scope operators. For the human architecture itself, there is no requirement that the various procedures available to the workspace be mutually consistent, only that the trouble caused by inconsistency be less than the inferential benefits gained under normal consistions. Task-switching and scope-limiting mechanisms also prevent the emergence of contradictions during ordinary functioning, which makes the mutual consistency of the architecture as an abstract formal system not relevant. Mental logic hypotheses for human reasoning have been rejected empirically by many on the assumption that the only licensed inferences are logical. We believe that the content sensitivity of human reasoning is driven by the existence of domain-specific inference engines, which coexist beside operators that parallel more traditional logical elements.

[6]As well as heterarchical relations, governed by rules for data incorporation from other sources.

(2) *The statement is false that:*
(3) *Anthropology is a science.*

In this case, the scope operator (2) binds the scope within which the information of the data structure (3) can be accessed, so that (3) is not free to be promoted to the ground state or to be used elsewhere in the system. In contrast, the function of an explicit true tag in a statement description operator (i.e., *The statement is true that p)* would be to release the statement from previous scope restriction, promoting it to the next left-most level, or, if it was originally only one level down, changing its status to unmarked, or architecturally true.[7] Time and location operators operate similarly:

(4) *In ≠Tobe* (!Kung for "autumn"),
(5) *the mongongo nuts become edible and plentiful.*

or

(6) *At Nyae Nyae,*
(7) *there are chrysomelid beetles suitable for making arrow poison.*

Scope operators define, regulate, or modify the relationships between sets of information, and the migration of information between levels. They involve a minimum of two levels, a superordinate (or ground) level and a subordinate level. In these cases, the subordinate propositions cannot be reproduced without their respective scope tags, which describe the boundary conditions under which the information is known to be accurate, and which therefore license their use in certain inferences, but not others. As with classical conditioning, we expect that additional mechanisms are designed to keep track of the reality of the scope boundaries; for example, observing a lack of contingency outside the boundaries may eventually release the restriction. Thus, (6–7) may be transformed into (7) for an individual whose travels from camp to camp are typically inside the beetle species' range. Conversely, architecturally true statements like (1) can be transformed by a scope operation into something scope limited, as new information about its boundary conditions are learned. A time-based scope transformation would be as follows:

(8) *It is no longer true that*
(9) *anthropology is a science.*

Scope operators regulate the migration of information into and out of subordinated data sets, coupling (allowing data to flow) and decoupling them according to the nature of the operator and the arguments it is fed. They bind propositions into internally transparent but externally regulated sets.

In so doing, they provide many of the tools necessary to solve the problems posed by contingent information. By imposing bounds on where scope-limited

[7]Promotion is equivalent to Tarskian disquotation, with respect to the next level in the architecture.

information can travel (or what can access it), it allows information to be retained by the system and used under well-specified conditions, without allowing it to damage other reliable data sets through inferential interaction. We will call representations that are bound or interrelated by scope operators *scope-representations* or *S-representations*.

Since computational features evolve because they enhance behavioral regulation, it is worth noting that these innovations markedly increase the range of possible behaviors open to the organism. In particular, one major change involves *acting as if*. The organism would be highly handicapped if it could only act on the basis of information known to be true, or have its conduct regulated by architecturally true propositions, although this was likely to be the ancestral state of the organism. With the ability *to act as if p*, or *to act on the basis of p*, the organism can use information to regulate its behavior without losing any scope-represented restrictions on the nature of the information, or without necessarily losing a continuing awareness that the information acted on is not or might not be true. Conditions where such a behavioral-representational subsystem are useful include the many categories of actions undertaken under conditions of uncertainty (e.g., *We will assume they got the message about the restaurant.*; or *We will act as if there is a leopard hiding in the shadows of the tree.*); actions with respect to social conventions or deontic commitments (which are by themselves incapable of being either true or not true, at least in an ordinary sense; e.g., *Elizabeth is the rightful Queen of England.*; *It is praiseworthy to make the correct temple sacrifices.*); adapting oneself to the wishes of others; hypothesis testing, and so on.[8] Pretense (Leslie, 1987) and deception (Whiten & Byrne, 1997) are simply extensions of this same competence, in which the agent knows the representations on which he or she is acting are false. (Deception and pretense are limiting cases in which the information is S-represented as false with 100% certainty. Typically, however, S-representations will be tagged with more intermediate credal values.) In order to get coordinated behavior among many individuals, and the benefits that arise from it, it is necessary to agree on a set of representations that will be jointly acted upon—a reason why social interaction so often involves the manufacture of socially constructed but unwarranted shared beliefs. Structures of representations can be built up that can be permanently consulted for actions, without their contents unrestrictedly contaminating other knowledge stores.

Credal values and modals (*it is likely that p; it is possible that p; it is certain that p*) allow the maintenance and transformation of scope-marked information bound

[8]Indeed, this kind of architecture offers a computational explanation of what kind of thing deontic ascriptions are: decoupled descriptions of possible actions and states of affairs, of suspended truth value, connected to value assignments of the possible actions.

to information about likelihood and possibility—regulatory information that often changes while the underlying propositions are conserved. Propositional attitude verbs (e.g., *think*, *believe*, *want*, *hope*, *deny*) are obviously a key category of scope operator as well (Leslie, 1987; Baron-Cohen, 1995).

Supposition, Counterfactuals, and Natural Deduction Systems

What makes such a system resemble, to a certain extent, natural deduction systems is the presence of scope operators, such as supposition, and the fact that these operators create subdomains or subordinate levels of representation, which may themselves have further subordinate levels, growing into multilevel, treelike structures. Supposition involves the introduction of propositions of unevaluated or suspended truth value, which are treated as true within a bounded scope and then used as additional content from which to combinatorially generate inferential products. The operator "if," for example, opens up a suppositional world (e.g., *I am in my office this afternoon. If students believe I am not in my office this afternoon, then they won't bother me. If I close my door, and leave my light off, they will believe I am not here.*) whose contents are kept isolated from other proposition sets, so that true propositions are not intermixed and hence confused with false ones (e.g., *I am not in my office*) or potentially false ones (e.g., *they won't bother me*). Any number of subordinate levels can be introduced, with additional subordinate suppositions or other scope operations.

A key feature of such a deduction system is the restricted application of inferences. Inferences are applied, in a rule-governed but unrestricted fashion within a level (e.g., *students believe I am not in my office this afternoon*, therefore, *they won't bother me*), but not across levels (e.g., there is no contradiction to be recognized between *I am in my office this afternoon* and the proposition *I am not in my office this afternoon* because they are at different levels in the structure; Leslie & Frith, 1990). Contents are architecturally true with respect to the level they are in and may enter into inferences at that level, while remaining false or unevaluated with respect to both the ground state of the architecture and other intermediate superordinate levels. Certain propositional attitudes (e.g., believe as opposed to know) also decouple the truth value of the propositions (*I am not in my office*) that are embedded in encompassing statements, a process that can be dissected computationally. Paradoxically, an architecture that only processes true information is highly limited it what it can infer, and most forms of human discovery by reasoning involve supposition. While some cases are famous, such as Newton's thought experiment in (*10*), normal cases of suppositions are so numerous that they permeate our thoughts in carrying out routine actions in our daily lives (*11*).

(10) Suppose I threw this rock hard enough that the earth fell away in its curvature faster than the rock's downward ballistic took it?

(11) What if I hold my airline ticket in my teeth while I pick up the baby with my right arm and our bags with my left arm?

Supposition (e.g., *12*) is a scope operation that suspends truth values for all successive computations that result from taking the supposition as a premise, which in this case is only (*13*).

(12) Suppose my wife, Desdemona, was unfaithful with Cassio.

(13) Then Cassio, who I thought was my friend, has betrayed me.

Suppositions and their entailments remain internally interrelated and generative but isolated from the rest of the data in the architecture. If (*13*) was allowed to escape its scope restriction to enter into ground state-originating inferences, the effects would be disastrous. Othello would have (*13*) as part of his uncritically accepted semantic store of propositions, without it being warranted (or "true" within the decoupled world of Shakespeare's *Othello*).[9] Nevertheless, S-representations like (*12–13*) allow many types of useful and revelatory reasoning to proceed—everything from proof by contradiction to the construction of contingency plans. Additionally, suppositions contain specifications of when subordinate deductions can be discharged. This occurs when other processes produce a true proposition that duplicates that supposition. Evidence establishing (*12*) as true discharges the supposition, promoting (*13*) to architectural truth and stripping it of its scope restrictions.

Actions can also discharge suppositions—a key point. Consider a hominid considering how to capture a colobus monkey in a tree. An architecture that cannot consider decoupled states of affairs is limited in the behaviors it can take (e.g., close distance with monkey). This may often fail because of the nature of the situation. Consider, for example, a situation in which a branch from the monkey's tree is close to the branch of a neighboring tree. In this situation, the hominid confronts the following contingencies: If he climbs the trunk, then the monkey escapes by the branch. If he climbs across the branches, then the monkey escapes by the trunk. Before taking action, if the hominid suppositionally explores the alternative hunt scenarios, then he will detect the prospective failure. Moreover, given alternative inferential pathways leading to failure, the hominid, armed with the inferential power of supposition (and various other inferential tools, such as a model of the prey mind and a theory of mechanics), may then begin to consider additional courses of action suppositionally, reasoning about the likely consequences of each alternative. *Suppose there were no branch on the neighboring tree, then it could not be used as an escape route. Suppose, before I initiate the hunt by climbing up the trunk, I break that branch. Then it could not be used as an escape route. If I then go up the trunk, the monkey cannot escape. The hunt will be a success. End search for successful outcome. Transform suppositional structure into a plan.* Conveniently for planning and action, the conditions for discharging a supposition specify the actions that need to be taken to put that aspect of the plan into effect, and the tree structure of suppositions provides the information about the order of the causal steps to be taken. Hominids armed with suppositional reasoning can undertake new types of successful behaviors that would be impossible for those whose cognitive architectures lacked such design features. It allows them to explore the properties of situations computationally in order to identify sequences of improvised behaviors that may lead to novel, successful outcomes. The restricted application of inferences to a level, until suppositions (or other scope limitations) are discharged, is a crucial element of such an architecture. The states of affairs under the scope of a specific supposition are not mistaken for states of affairs outside that supposition: superordinate and subordinate relationships are kept clear until their preconditions can be discharged (as when an action is taken).

Like a clutch in an automobile, supposition and other scope operators allow the controlled engagement or disengagement of powerful sets of representations, which can contain rich descriptions, with domain-specific inference engines, which can be applied when their preconditions are met. These operators provide vehicles whereby information which may or may not be counterfactual can be processed without the output being tagged as *true* and stored as such.

Because contingent information can change its status at any time with any new change in the world, it is important to have tools available that can take architecturally true information and scrutinize it. For example, the workspace that contains proposition *p* may benefit from demoting *p* into the scope-representation, *It appears that p*. Proposition *p* can still provide the basis for action but can now be subjected to inferential processes not possible when it was simply a free representation at ground state. Demotion into a scope-representation brings a representation out of architectural truth and into a new relationship with the primary workspace.[10] Because of this feature of the human cognitive architecture, humans can contingently refrain from being naïve realists about any specific data structure, although presumably we will always be naïve realists about whatever happens to be in the ground state in the workspace at any given time.[11]

[9]Such an architecture explains how humans process fictional worlds without confusing their environments and inhabitants with the real world.

[10]It is interesting in this regard that children's ability to distinguish appearance from reality (e.g., on seeing a sponge that looks like a rock, they can say that although it looks like a rock, it is really a sponge) matures at about the same time as their ability to represent false beliefs (about 4 years). This is sometimes interpreted as evidence for the maturation of a theory of mind (Baron-Cohen, 1995). It might, however, reflect the maturation of a more encompassing scope syntax, of which M-representations are a part.

[11]We think that ground-state representations are present in consciousness but are not automatically the objects of consciousness; that is, we are not automatically reflectively conscious of these data structures, although they can easily be made so. Data structures in the ground state must be demoted to become the object of inferential scrutiny. Indeed, we think that the function of the architectural component that corresponds to one referent of the word *consciousness* is to be a buffer to hold isolated from the rest of the architecture the intermediate computational work products during the period when their truth value and other merits are unevaluated. This explains why consciousness is so notoriously volatile.

Some operators are recursive, and some types of subordinated data structures can serve as the ground for further subordinated structures, leading potentially to a tree structure of subordinated and parallel relations whose length and branching contingencies are restricted only by performance limitations of the system. For example:

(14) *Anna was under the impression that*
(15) *Clifford has claimed that*
(16) *most anthropologists believe that*
(17) *the statement is false that:*
(18) *anthropology is a science.* [and]
(19) *quantum physicists have demonstrated that:*
(20) *science is only an observer-dependent set of arbitrary subjective opinions.*

Extensive thinking about a topic can produce structures too elaborate to be placed, in their entirety, into the workspace, and which are therefore considered in pieces. The cultural development of memory aids such as writing have allowed an explosion of conceptual structures that are larger than what our ancestors would have routinely used.

Scope operators greatly augment the computational power of the human cognitive architecture, compared with ancestral systems lacking such features. One advantage of an architecture equipped with scope operators is that it can carry out inferential operations on systems of inferences of unevaluated or suspended truth value, while keeping their computational products isolated from other knowledge stores until the truth or utility of the elements can be decided. It they were not kept isolated, their contents would enter into inferences with other data structures in the architecture, often producing dangerously false but unmarked conclusions (e.g., *science is only a set of arbitrary subjective opinions* would be disastrous guidance for someone who has to choose a medical strategy to arrest an epidemic in a developing country). Fortunately, (*14*) decouples the uncertain information in (*15–20*) from the rest of the architecture but allows the information to be maintained, and reasoned about, within various lawful and useful restrictions specified in the scope operators. The structure (*14–20*) is free to migrate through the system as a bound unit, entering into whatever licensed inferences it can be related to, but its subordinate elements are not.

Within subordinate levels (*15–20*), similar scope operations structure the inferences that are possible. The operator *demonstrate* assigns the value "true" to the subordinate element (*20: science is only*...), allowing its contents to be promoted to the next level. Within that level, it is treated as true, although it is not true above that level or outside of its scope-circumscription. The operator that governs that level—*claim*—prevents it from migrating independently of the metarepresentation it is bound to (*Clifford has claimed that*...). Both (*16*) plus entailments and (*19*) plus entailments are true within the world of Clifford's claims and are free to inferentially interact with each other, along with (*20*) and any other of Clifford's claims that turn up. Indeed, one can say that a representation is true with respect to a particular level in a particular data structure; any level can function as a ground level to subordinate levels. It is scope-conditionally true for a data structure when it is permitted by the architecture to interact with any other information held within the same or subordinate levels of that data structure.

Source, Error Correction, and the Evolution of Communication

Different scope operators obviously have different regulatory properties and, hence, different functions. *Claim, believe,* and *demonstrate,* for example, require source tags as arguments, as well as conveying additional information (e.g., publicly assert as true that *p*; privately treat as architecturally true that *p*; and have publicly established the truth that *p*, respectively).[12] Source tags are very useful because often, with contingent information, one may not have direct evidence about its truth but may acquire information about the reliability of a source. If the sources of pieces of information are maintained with the information, then subsequent information about the source can be used to change the assigned truth status of the information either upwards or downwards. For example, one may not assign much credal value to what most anthropologists believe (*16*), or one may discover that Clifford in particular is highly unreliable (*15*), while having a solid set of precedents in which Anna's impressions (such as *14*) have proven highly reliable, despite the fact that she herself is unwilling to evaluate her impressions as trustworthy. Sources may include not only people but also sources internal to the architecture, such as vision, episodic memory, a supposition, previous inference, and so on. Thus, humans can have the thought "My eyes are telling me one thing, while my reason is telling me another."

In general, our minds are full of conclusions without our having maintained the grounds or evidence that led us to think of them as true. For a massively inferential architecture like the human mind, each item can serve as input to many other inferential processes, whose outputs are inputs to others. To the extent that the information is sourced, or its grounds and derivation are preserved in association with the data, then new data about its evidential basis can be used to correct or update its inferential descendants. (If "Stock in Yahoo is a good investment" is tagged with its source—Gordon Gekko—it can be reevaluated when you learn that Gekko has been indicted for insider trading and stock manipulation.) To the extent that information is not sourced, or its process of infer-

[12]We are not claiming that every propositional attitude term, for example, is reliably developing or innate. We consider it more plausible that there is an evolved set of information-regulatory primitives that can be combined to produce a large set of scope operators and scope representations.

ential derivation is not preserved in association with it, then it cannot be automatically corrected when the grounds for belief are corrected. (Even more basically, Sperber has persuasively argued that the inferential nature of communication itself requires the online metarepresentational processing of language in order for interpretation to be successful [Sperber & Wilson, 1986; Sperber 1985, 1996, 2000].)

Indeed, our minds are undoubtedly full of erroneous inferential products that were not corrected when their parent source information was updated because they could no longer be connected with their derivation. Because source tags, and especially pathways of derivation, are costly to maintain, mechanisms should monitor for sufficient corroboration, consistency with architecturally true information, or certification by a trusted source. If, or when, a threshold is reached, the system should no longer expend resources to maintain source information, and it should fade. This is what makes trust so useful (one does not need to keep the cognitive overhead of scope-processing communication) but so dangerous (one cannot recover and correct all of the implanted misinformation). After all, what is important about an encyclopedia of (accurate) knowledge about the world is the facts themselves; not who told them to you, what their attitude toward them was, or when you learned them. Typically, once a fact is established to a sufficient degree of certainty, source, attitude, and time tags are lost (Sperber, 1985; Tulving, 1983; Shimamura, 1995); for example, most people cannot remember who told them that apples are edible or that plants photosynthesize. Moreover, an encyclopedia is most useful when the facts can cross-reference one another, so that each can support inferences that may apply to others, thereby adding further, inferred facts to the body of knowledge (e.g., *Mercury is a poison.*; *Tuna has high levels of mercury.*; therefore, *People who eat tuna are ingesting poison.*). This means that truth conditions must not be suspended for facts in semantic memory, and the scope of application for any truth-preserving inference procedures must be relatively unrestricted within the encyclopedia, such that facts can mate promiscuously to produce new, inferred facts.

CONCLUSION

Since Frege, philosophers have been aware that propositional attitudes suspend semantic relations such as truth, reference, and existence (Frege, 1892; Kripke, 1979; Richard, 1990). Frege noticed, for example, that the principle of substitution of coreferring terms breaks down when they are embedded in propositional attitudes (i.e., one can believe that Batman fights crime without believing that Bruce Wayne fights crime). Or, consider the following statement:

(25) Shirley MacLaine believes that

(26) she is the reincarnation of an Egyptian princess named Nefu.

(25–26) can be true, without Nefu ever having existed, and without it being true that Shirley is her reincarnation. The propositional attitude verb *believe* suspends truth, reference, and existence in (26), fortunately decoupling (26) from the semantic memory of those who entertain this statement.

Rather than being quirks, problems, and puzzles, as philosophers have often regarded them, it seems possible that such suspensions are instead adaptations: design features of a computational architecture designed to solve problems posed by the many varieties of contingent information exploited by our ancestors, and the interrelationships among sets of contingent information. To benefit from contingent information without being destroyed by it, the human cognitive achitecture must be equipped with a scope syntax. It seems likely that scope-representations and operators are reliably developing, species-typical features of the human cognitive architecture, and that design features of this kind are necessary—though not sufficient—for any system that manifests improvisational intelligence.

In sum, the picture of intelligence that emerges from the collision of evolutionary biology and cognitive science differs in many ways from more commonly held conceptions of what intelligence consists of. Such an evolutionary analysis throws doubt on some views (i.e., intelligence as a set of content-independent rational methods), necessitates some distinctions (i.e., between dedicated and improvisional intelligence), and appears to solve some questions (Why is improvisional intelligence so zoologically rare?). Nevertheless, it also uncovers a further set of questions (i.e., How can computational procedures evolve that exploit the novel features of unique situations?) that deepen the enigma of human intelligence and indicate that building an accurate model of the computational machinery underlying human intelligence will require novel insights that only improvisional intelligence can supply.

REFERENCES

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month old infants. *Cognition, 23*, 21–41.

Barkow, J., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Boyd, R., & Richerson, P. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Boyer, P. (1993). *The naturalness of religious ideas*. Berkeley: University of California Press.

Burke, T., Davies, N., Bruford, M., & Hatchwell, B. (1989). Parental care and mating behaviour of polyandrous dunnocks *Prunella modularis* related to paternity by DNA fingerprinting. *Nature, 338*, 249–251.

Clark, A. (1997). *Being there*. Cambridge, MA: MIT Press.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187–276.

Cosmides, L., & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupre (Ed.), *The latest on the best: Essays on evolution and optimality*. Cambridge, MA: The MIT Press.

Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, Part II. Case study: A computational theory of social exchange. *Ethology and Sociobiology, 10*, 51–97.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

Cosmides, L., & Tooby, J. (1996a). A logical design for the mind? *Contemporary Psychology, 41*(5), 448–450.

Cosmides, L., & Tooby, J. (1996b). Are humans good intuitive statisticians after all? Rethinking some conclusions of the literature on judgment under uncertainty. *Cognition, 58*, 1–73.

Cosmides, L., & Tooby, J. (1997). Dissecting the computational architecture of social inference mechanisms. In *Characterizing human psychological adaptations*. Ciba Foundation Symposium: Vol. 208. Chichester, UK: Wiley.

Cosmides, L., & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentations. In D. Sperber (Ed.), *Metarepresentation*. Vancouver Studies in Cognitive Science. New York: Oxford University Press.

Davies, N. (1989). The dunnock: Cooperation and conflict among males and females in a variable mating system. In P. Stacey & W. Koenig (Eds.), *Cooperative breeding in birds* (pp. 457–485). Cambridge: Cambridge University Press.

Dawkins, R. (1976). *The selfish gene*. New York: Oxford University Press.

Dawkins, R. (1982). *The extended phenotype*. San Francisco: W.H. Freeman.

Dawkins, R. (1986). *The blind watchmaker*. New York: Norton.

Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

de Waal, F. (1982). *Chimpanzee politics: Power and sex among apes*. New York: Harper & Row.

Frege, G. (1892). On sense and reference. In P. Geach & M. Black (Eds.), *Translations of the philosophical writings of Gottlob Frege*. Oxford, England: Blackwell.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

Garcia, J. 1990. Learning without memory. *Journal of Cognitive Neuroscience, 2*, 287–305.

Gendron, R. P., & O. J. Reichman. (1995). Food perishability and inventory management: A comparison of three caching strategies. *American Naturalist 145*, 948–968.

Gentzen, G. (1969). Investigations into logical deduction. In M. E. Szabo (Ed.), *The collected papers of Gerhard Gentzen* (pp. 405–431). (Original work published 1935). Amsterdam: North Holland Publishing Co.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.

Gigerenzer, G., Todd, P., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gould, S., & Lewontin, R. (1979). The spandrels of San Marco and the Panglossian program: A critique of the adaptationist program. *Proceedings of the Royal Society of London, 250*, 281–288.

Heinrich, B. (1981). *Bumblebee economics*. Cambridge, MA: Harvard University Press.

Hirschfeld, L., & Gelman, S. (Eds.). (1994). *Mapping the mind: Domain-specificity in cognition and culture*. New York: Cambridge University Press.

Humphrey, N. (1984). *Consciousness regained*. Oxford: Oxford University Press.

James, W. (1892). *The principles of psychology*. London: Macmillan.

Johnson, M., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Oxford, England: Blackwell.

Kahneman, D., Slovic, P., & Tversky, A. (Eds). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kalat, J., & Rozin, P. (1973). "Learned Safety" as a mechanism in long-delay taste-aversion learning in rats. *Journal of Comparative & Physiological Psychology, 83*, 198–207.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind reading and manipulation. In Krebs, J. R., & Davies, N. B. (Eds.). *Behavioural ecology: An evolutionary approach* (2nd ed., pp. 380–402). Oxford, England: Blackwell.

Kripke, S. (1979). A puzzle about belief. In A. Margalit (Ed.) *Meaning and use*. (pp. 239–283). Dordrecht, Holland: Reidel.

Lee, R. B. (1993). *The Dobe Ju/'hoansi.* (2nd ed.). New York: Holt, Reinhart & Winston.

Leslie, A. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review, 94*, 412–426.

Leslie, A. (1988). The necessity of illusion: Perception and thought in infancy. In L Weiskrantz (Ed.), *Thought without language* (pp. 185–210). Oxford, UK: Clarendon Press.

Leslie, A. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Leslie, A., & Frith, U. (1990). Prospects for a cognitive neuropsychology of autism: Hobson's choice. *Psychological Review, 97*, 122–131.

Mandler, J. M., & McDonough, L. (in press). Studies in inductive inference in infancy. *Cognitive Psychology*.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Mineka, S., & Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *Journal of Experimental Psychology: General, 122*, 23–38.

Mithen, S. (1996). *The prehistory of the mind*. London: Thames & Hudson.

Nozick, R. (1993). *The nature of rationality*. Cambridge, MA: Harvard University Press.

Ohman, A., Dimberg, U., & Ost, L. G. (1985). Biological constraints on the fear response. In S. Reiss & R. Bootsin (Eds.), *Theoretical issues in behavior therapy* (pp. 123–175). New York: Academic Press.

Perrigo, G., Bryant, W., & vom Saal, F. (1990). A unique neural timing mechanism prevents male mice from harming their own offspring. *Animal Behavior, 39*, 535–539.

Pinker, S. (1994). *The language instinct*. New York: Morrow.

Pinker, S. (1997). *How the mind works*. New York: Norton.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences 13*, 707–784.

Post, D., McDonald, M., & Reichman, O. (1998). Influence of maternal diet and perishability on caching and consumption behavior of juvenile easter woodrats. *Journal of Mammalogy, 79*, 56–162.

Premack, D., & Premack, A. (1997). Infants attribute value+- to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience, 9*, 848–856.

Pressley, P. (1981). Parental effort and the evolution of nest-guarding tactics in the ghreespine stickleback, *Gasterosteus aculeatus L. Evolution, 35*, 282–295.

Richard, M. (1990). *Propositional attitudes: An essay on thoughts and how we ascribe them*. Cambridge, MA: Cambridge University Press.

Rips, L. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA.: MIT Press.

Rozin, P., & Kalat, J. (1971). Specific hungers and poison avoidance as adaptive specializations of learning. *Psychological Review, 78*, 459–486.

Seligman, M. (1971). Phobias and preparedness. *Behavior Therapy, 2*, 307–320.

Shepard, R. N. (1992) The three-dimensionality of color: An evolutionary accommodation to an enduring property of the world? In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Shimamura, A. (1995). Memory and frontal lobe function. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 803–813). Cambridge, MA: MIT Press.
Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14,* 29–56.
Sperber, D. (1985). Anthropology and psychology: Towards and epidemiology of representations. *Man, 20,* 73–89.
Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 39–67). New York: Cambridge University Press.
Sperber, D. (1996). *Explaining culture: A naturalistic approach.* Oxford, England: Blackwell.
Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentation.* Vancouver Studies in Cognitive Science. New York: Oxford University Press.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Oxford, England: Blackwell.
Symons, D. (1993). The stuff that dreams aren't made of: Why wake-state and dream-state sensory experiences differ. *Cognition, 47,* 181–217.
Tager-Flusberg, H., Boshart, J., & Baron-Cohen, S. (1998). Reading the windows to the soul: Evidence of domain-specific sparing in Williams syndrome. *Journal of Cognitive Neuroscience, 10,* 631–639.
Tooby, J., & Cosmides, L. (1989). Evolutionary psychology and the generation of culture, Part I: Theoretical considerations. *Ethology & Sociobiology, 10,* 29–49.
Tooby, J., & Cosmides, L. (1990). The past explains the present: emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology, 11,* 375–424.
Tooby, J., & Cosmides, L. (1992a). *Ecological rationality and the multimodular mind: Grounding normative theories in adaptive problems.* (Center for Evolutionary Psychology Technical Report 92–1). Center for Evolutionary Psychology, University of California, Santa Barbara.
Tooby, J., & Cosmides, L. (1992b). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press.
Tooby J., & DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. In W. Kinzey (Ed.), *Primate models of hominid behavior* (pp. 183–237). New York: SUNY Press.
Tulving, E. (1983). *Elements of episodic memory.* Oxford: Oxford University Press.
Turner, M. (1996). *The literary mind.* New York: Oxford University Press.
Uexkull, J. von. (1957). A stroll through the world of animals and men: A picture book of invisible worlds. In C. H. Schiller (Ed. & Trans.), *Instinctive Behavior: The development of a modern concept,* (pp. 5–80). New York: International Universities Press.
Wehner, R., & Srinivasan, M. (1981). Searching behavior of desert ants, genus *Cataglyphis* (*Formicidae*, Hymenoptera). *Journal of Comparative Physiology, 142,* 315–338
Whiten, A., & Byrne, R. (1997). *Machiavellian intelligence II: Extensions and evaluations.* Cambridge: Cambridge University Press
Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought.* Princeton, NJ: Princeton University Press.